THE UNIVERSITY
of EDINBURGH

Exploring and Expanding the Synthetic Potential of a Biocatalytic Dynamic Kinetic Resolution

Silvia De Cesare

A Thesis Submitted for the Degree of Doctor of Philosophy

The University of Edinburgh
2020

Lay Summary

Enantiopure canonical and non-canonical amino acids (NCAA) have found numerous applications over the years as chiral building blocks employed in the pharmaceutical, agrochemical and fine chemical industries and as fundamental tools for enzyme evolution and development of biotherapeutics *via* genetic code expansion. Unfortunately, the classic chemosynthetic strategies for the production of these compounds relies on the use of toxic reagents and usually lead to the formation of a racemic mixture. In recent years numerous environmentally friendly biocatalytic strategies have been developed to achieve the stereoselective synthesis of NCAAs in high yields and enantiomeric excesses. The employment of enzymes in synthesis brings many advantages as the use of biodegradable catalysts, mild reaction conditions and intrinsic high selectivity, however biocatalysis is also characterised by various limitations, such as low protein stability, unfavourable reaction equilibrium and enzyme's narrow substrate scope. Strategies as enzyme engineering and immobilization have been developed and employed to try to overcome the biocatalysts limitations.

In this work the coupling of an engineered *N*-acetyl amino acid racemase (NAAAR) with two deacetylases: *E. coli* ArgE and *S. viridochromogenes* Tü494 Dea, will be explored to develop a biocatalytic dynamic kinetic resolution for the production of the natural herbicide phosphinothricin and other highly valuable NCAAs.

Abstract

The market demand for enantiopure non-canonical amino acids (NCAA) has been increasing in recent years, given the importance of these compounds as chiral building blocks employed in the synthesis of many pharmaceuticals and agrochemicals. An example is the use of L-phosphinothricin (L-PT), a bioactive amino acid with potent herbicidal properties produced as a secondary metabolite by the Gram positive bacterium *Streptomyces viridochromogenes*. Recently, the use of NCAAs has been expanded to include applications in protein engineering, thanks to the advances achieved in the field of genetic code expansion. Over the last decade various biocatalytic strategies have been developed for the enantioselective production of these compounds, using a variety of enzymes, such as aminotransferases and ammonia lyases. A well-established biocatalytic protocol allows for the efficient preparation of NCAAs from their *N*-acylated derivatives *via* a dynamic kinetic resolution (DKR). This is achieved by coupling an engineered *Amycolatopsis N*-acyl amino acid racemase (NAAAR) with a compatible stereoselective deacetylase (amide hydrolase).

Two recombinant hydrolytic enzymes: *E. coli* ArgE and *S. viridochromogenes* Tü494 Dea have been investigated in this work. Purified ArgE is a zinc-dependant, L-selective *N*-acetyl-L-ornithine deacetylase and a thorough kinetic study revealed that it displays a wide substrate scope but poor activity for acidic and aromatic substrates. Various tests were carried out to optimise biotransformation conditions, which were employed in the coupling of ArgE with the engineered NAAAR double mutant (NAAAR DM: G291D F323Y) for the DKR of *N*-acetyl NCAAs derivatives. Directed evolution was employed to increase deacetylase activity towards acidic *N*-acetylated amino acids; unfortunately, thus far no improved variant has been isolated. Prism-shaped protein crystals were obtained, but with poor resolution, which prevents the determination of the *E. coli* ArgE X-ray structure. Despite this, homology modelling and site-directed mutagenesis were successfully employed to probe the enzyme active and substrate binding sites.

The deacetylase *S. viridochromogenes* Dea is a novel member of the hormone-sensitive lipase (HSL) family. It was proposed that this enzyme catalyses the hydrolysis of *N*-Ac-L-bialaphos, the precursor of the tripeptide bialaphos (PTT), a natural herbicide that contains the L-PT building block. Similar to other members of the HSL family *Sv*Dea catalyses the deacetylation reaction by employing a catalytic triad of highly conserved serine, histidine and glutamate

residues in the active site, but the basis of its substrate specificity is unknown. To study its synthetic utility recombinant *Sv*Dea was isolated from *E. coli* and activity screening using various assays (including a novel ^{1}H NMR assay) revealed a narrow substrate scope, limited to small acetylated esters and *N*-acetyl-L-PTT. Initial crystal trials failed to deliver positive hits, so structure and sequence analysis generated a homology model to obtain further insights into the substrate binding and direct future engineering efforts.

In the last section of this thesis the synthetic scope of a novel NAAAR quadruple mutant (NAAAR QM: Q26I M50I G291D F323Y) was explored. This engineered racemase was found to display enhanced activity for bulky *N*-acetylated amino acids and could be coupled with either L- and D-selective acylases to prepare enantiopure target L- or D- amino acids. Following small scale screening, eight DKRs were set up at 1 g scale to highlight the biocatalytic potential of this versatile enzyme for the preparation of valuable, optically pure phenylalanine derivatives from racemic starting materials.

Dedication

This work is dedicated to my mom and dad and to my grandmothers, Ester and Maria. For teaching me that through determination everything is possible in life and for giving me the love and support needed to achieve my goals.

“How would you like to live in Looking-Glass House, Kitty? I wonder if they'd give you milk in there? Perhaps Looking-Glass milk isn't good to drink?”

Lewis Carroll, ***Through the Looking-Glass***

“The aim of science is not to open the door to infinite wisdom, but to set a limit to infinite error [...] Nowadays, anyone who wishes to combat lies and ignorance and to write the truth must overcome at least five difficulties. He must have the courage to write the truth when truth is everywhere opposed; the keenness to recognize it, although it is everywhere concealed; the skill to manipulate it as a weapon; the judgment to select those in whose hands it will be effective; and the running to spread the truth among such persons.”

Bertolt Brecht, ***Life of Galileo***

“È di nuovo fra noi, in un bicchiere di late. È inserito in una lunga catena, molto complessa, tuttavia tale che quasi tutti i suoi anelli sono accetti nel corpo umano. Viene ingoiato: e poiché ogni struttura vivente alberga una selvaggia diffidenza verso ogni apporto di altro materiale di origine vivente, la catena viene meticolosamente frantumata, ed i frantumi, uno per uno, accettati o respinti. Uno, quello che ci sta a cuore, varca la soglia intestinale ed entra nel torrente sanguigno: migra, bussa alla porta di una cellula nervosa, entra e soppianta un altro carbonio che ne faceva parte. Questa cellula appartiene ad un cervello, e questo è il mio cervello, di me che scrivo, e la cellula in questione, ed in essa l'atomo in questione, è addetta al mio scrivere, in un gigantesco minuscolo gioco che nessuno ha ancora descritto. È quella che in questo istante, fuori da un labirintico intreccio di sì e di no, fa sì che la mia mano corra in un certo cammino sulla carta, la segni di queste volute che sono segni; un doppio scatto, in su ed in giù, fra due livelli d'energia guida questa mia mano ad imprimere sulla carta questo punto: questo.”

Primo Levi, ***Il Sistema Periodico***

Acknowledgments

First of all, I would like to thank Prof. Dominic Campopiano for giving me the opportunity to join his group and for his unwavering support and patience. He has been a great supervisor, showing never-ending patience, moral support and useful suggestions during the ups and downs of my PhD. I also would like to thanks the Syngenta R&D team in Bracknell, especially Dr. Nick Mulholland and Dr. Lauren Ray for supervising me during my three-month industrial placement. They have been very helpful and I learned so much during my time with them.

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) through the CRITICAT doctoral training programme which has not only allowed me to develop my scientific skills but also meet some incredible people. I wish to thank Dr Kevin Jones for organising all the CDT meetings, conferences and workshops and just to be a nice guy, always available to help out. I am also grateful to Dr. Anibal Cuetos and Prof. Gideon Grogan (University of York) for all their help in carrying out crystal trials. I would also like to thank Dr. Logan Mackay, Dr. Alan Taylor, Dr. Faye Cruickshank, Dr. Lorna Murray, Dr Juraj Bella and Dr. Lorna Eades for their help and trying to keep the MS, NMR and ICP facilities working. Also, a big thanks to all the technician who keep the School of Chemistry and our equipment running.

My thanks go to all the members of Lab 229, without whom my four years in Edinburgh would not have been the same. It was an honour hosting dinner parties and baking for you guys! In particular thank you to Cath, for always be available to be distracted by my chit-chat, all the organic synthesis advices and for being there whenever I was feeling down, your support has kept me going during the last four years. Also, a special recognition to Shona, for being a great colleague and an amazing friend to hang out, especially for afternoon teas; I promise we will manage to go to Italy together. I will miss our Dream Team chats, coffee breaks, lunches, runs and Fiit sections. Thanks to Gary for sharing his cloning/molecular biology knowledge with me and also his passion for puppies, video games and cooking. Thanks to Alexis for being the best crazy French pal I would ever meet, with our discussion on French wines and Italian cheese you made the last four years fly, also thanks for teaching me how to use the ATKA. Whenever you guys feel like having nice food, or chilling at the beach, I will be waiting in Italy, my door is always open. Thanks to Ben for the great banter and for letting me realise I lack a talent for singing at least for karaoke, and helping in keeping the ATKAs

functional. Thanks to Jo for teaching me MS and being a great person with a huge love for cats. Thanks to Richard for sharing his knowledge with me and all the patience he showed whenever I surprised him popping-up in his lab or office with 100 questions; he contributed a lot to my knowledge pool. You still can't hide from me, but next time I will seek you out just to deliver a thank you cake. Thanks to Pete, Annabel and Piera for introducing me to the lab and thanks to Alice, Michael, Rhona and Zenam for making the lab a lively and nice place to work. Also, a special acknowledgement to my students: Corrie, Young-Joo, Gabriel and Carolina, who have significantly contributed through their hard work to both this thesis and other side-projects. It has been a pleasure to meet you all, hopefully my teachings were not terribly boring. A final thanks goes to Wikipedia for helping me to understand all the biological jargon, when I first started my journey in the world of biocatalysis.

A thank you to all my friends for your undying support. In particular thanks to Iva, who made my six months spent in St Andrews so special, best of luck on your new job. Also, a thanks to Briony, for just being an amazing person to hung-up and team during CRITICAT workshops. I also want to thanks my flatmates Naoimh, Eva and Agis, for the nice dinners, gossip, painting critics and literature and theatre discussions and for teaching me the basics of weaving (or trying to), without you guys life would have been boring.

E finalmente siamo alla sezione italiana. Prima di tutto devo ringraziare i miei genitori per il loro infinito sostegno e per aver sempre creduto in me. La decisione di partire dall'Italia non è stata facile, ma nonostante qualche lacrima all'aeroporto, mamma e papà mi hanno sempre supportata e continueranno a farlo, specialmente ora che un nuovo viaggio sta per iniziare. Per favore, continuate ad inviarmi i sostentamenti, non posso sopravvivere nell'UK senza "i pacchi da giù". Un grande abbraccio va anche ai miei zii Maurizio, Stefano e alla zia Simo, venite a visitarmi non si sta mica così male qui, basta avere un ombrello ed una giacca per contrattaccare il clima britannico. Un ringraziamento speciale va anche agli zii e cugini di Genova (Gianna, Ito, Paola, Bunny, Fili e Pietro), di San Giuliano (Osanna Cristrina, Mauri, Ele e Massi) e di Valenza (Ivana, Gian Carlo, Marco, Ilari e la piccola Kiara) per tutti i pranzi, le cene, le festività, le serate su YouTube e le vacanze passate assieme. Se avrete mai voglia di visitare la "soleggiante" Scozia o Irlanda, fatemi un fischio. Non posso mancare di ringraziare la mitica professoressa Semino, o Nadia come preferisce essere chiamata adesso, se sono arrivata a prendere un PhD in chimica lo devo solo a lei, che per prima mi ha fatto scoprire una innata passione per la chimica e capire di avere la stoffa per diventare una scienziata.

Grazie ad Antonella per tutte le serate posh passate al teatro o a discutere l'ultimo libro letto (il principe Andrej sara' sempre nel mio cuore), in te ho trovato una una vera amica con cui condividere le mie passioni per l'opera e la letteratura russa e francese. E grazie a Gianluca per tutte le nostre pause caffè passate al dipartimento e le serate, pranzi ed aperitivi pieni di risate ed divertenti aneddoti sui vostri viaggi. Spero di essere stata almeno una buona distrazione dal lavoro, cosa che Nicole non mi perdonerà mai. Per me siete come il fratello e la sorella maggiore che non ho mai avuto e per questo sarete sempre i benvenuti a casa mia, vi procurerò una scorta a vita di Barbera. Grazie a Chiara per essere stata sempre al mio fianco in questi ultimi quattro anni, prima a Dundee come coinquilina e poi ad Edimburgo come amica in tutto e per tutto. Ogni momento passato insieme (dalle pause caffè ai pomeraggi di shopping) è stato prezioso ed indimenticabile. Grazie a Davide per avermi accolta in casa appena atterrata in città e la grande ospitalità dimostrata. Se ad oggi riesco a correrere 10 Km senza morire lo devo a sto ragazzo che mi ha ispirato a dare il meglio di me. Dobbiamo assolutamente correre assieme una volta, ti aspetto per la prossima mezza maratona. Ed è sempre grazie a sti due se sto spendendo troppe ore a guardare anime nei weekend. Voi avete reso speciale questa mia esperienza Edimburghese, vi aspetto tutti e quattro a Belfast per la prossima music night a suon di De Andre'.

Grazie a Martina e Margherita, per sempre coinquiline nel cuore, e le migliori compagne di viaggio (prossime mete New York e Tokyo), o inviate speciali da zone di guerra (#maiunagioa!) che una ragazza possa desiderare. In particolare grazie a Martina per i vari rants su Harry Potter e gli spoiler su Game of Thrones ed a Margherita per le sue imprese culinarie, come dimenticare il caffè carcerogeno ed il pollo Isis. Indipendentemente dalla distanza ci saro' sempre per voi, siete in fondo le mie due sorelle. Un grazie anche al grande Cattin ed al signor Boggio per le invettive politiche al cravino, i pranzi al B2, i pomeriggi passati tra retrosintesi e meccanismi di reazioni e le serate Pavesi in compagnia.

E per chiudere il mio sproloquio in bellezza un grazie speciale alle mie migliori amiche Alice, Alessia, Clarissa e Roxana, queste ragazze mi hanno accompagnata ad ogni passo ed erano sempre presenti, negli alti e nei bassi, a tutti i miei traguardi più importanti. Voi sarete sempre una parte fondamentale della mia vita; anche se credo che vederci tutte insieme possa causare gravi pericoli come ribellioni in Barcellona o pandemie globali, quindi e' meglio che il mondo si prepari per il nostro prossimo raduno.

Declaration

I, Silvia De Cesare, declare that the work presented is entirely my own and that it has not been submitted, in whole or in part, in any previous publication or application for a degree or professional qualification. Except where otherwise stated by reference or acknowledgment

Silvia De Cesare

The University of Edinburgh

2020

List of abbreviations

AA	Amino Acid
AADH	Amino acid dehydrogenase
AAR	Amino acid racemase
Abs	Absorbance
ACN	Acetonitrile
ADC	Antibody-drug conjugate
ADH	Alcohol dehydrogenase
ADP	Adenosine 5'-Diphosphate
AL	Ammonia Lyase
Alg	Allyl-glycine
AM	Amino mutase
API	Active pharmaceutical ingredient
ATP	Adenosine 5'-Triphosphate
BVMO	Bayer-Villager mono-oxygenase
CAPS	*N*-cyclohexyl-3-aminopropanesulfonic acid
Cit	Citrulline
Conv.	Conversion
CV	Column Volume
DAPDH	di-amino pimarate dehydrogenase
Dea	Deacetylase
DERA	Deoxyribose 5-phosphate aldolase
DH	Dehydrogenase
DKR	Dynamic Kinetic Resolution
DM	Double mutant
DMF	di-methylformammide
DMPT	de-methyl PT
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNFB	1,3-difluoro-4,6-dinitro benzene

dNTP	Deoxynucleotide Triphosphate
E. coli	*Escherichia Coli*
EDTA	Ethylene Diamine Triacetic Acid
EDXRF	Energy Dispersive X-Ray Fluorescence
ee	Enantiomeric Excess
EP-PCR	Error Prone-PCR
ESI	Electrospray Ionisation
EWG	Electron Withdrawing Group
FA	Formic Acid
FAD	Flavin adenine dinucleotide
$FADH_2$	Flavin adenine dinucleotide reduced form
FDAA	1-fluoro-2, 4-dinitrophenyl-5-L-alanine amide (Marfey's reagent)
FMOC	Fluorenylmethyloxycarbonyl
GF	Gel Filtration
GluDH	Glutamate dehydrogenase
GOase	Phosphopentomutase
GS	Glutamine synthase
h	hour
HF-PCR	High Fidelity-PCR
HPLC	High Performance Liquid Chromatography
HRP	Horseradish peroxidase
ICP	Inductively Coupled Plasma
IPA	Isopropylamine
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
k_{cat}	Catalytic Constant
K_d	Dissociation Constant
K_M	Michaelis-Menten Constant
KR	Kinetic Resolution
KRED	Ketoreductase
L-AAO	L-Amino Acid Oxidase
LB	Luria Bertani

LC	Liquid chromathography
LC-MS	Liquid Chromatography – Mass Spectrometry
LMW	Low Molecular Weight
LOD	Limit Of Detection
MAL	Methylaspartate ammonia lyase
MeOH	Methanol
min	minutes
MLE	Muconate lactonizing enzyme
MR	Marfey's reagent
MS	Mass Spectrometry
NAAAR	*N*-Acetyl Amino Acid Racemase
N-Ac	*N*-acetyl
NaCi	Sodium carbonate buffer
NAD^+	Nicotinamide Adenine Dinucleotide
NADH	Nicotinamide Adenine Dinucleotide Reduced Form
$NADP^+$	Nicotinamide Adenine Dinucleotide Phosphate
NADPH	Nicotinamide Adenine Dinucleotide Phosphate Reduced Form
NAM	*N*-acetyl methionine
NaPi	Sodium Phosphate buffer
NCAA	Non-canonical amino acid
NMR	Nuclear Magnetic Resonance
NMS	*N*-succinyl methionine
NRPS	Non-ribosomal peptide synthase
NRPS	Non-Ribosomal Peptide Synthetase
NSAAR	*N*-Succinyl Amino Acid Racemase
NSAT	*N*-succinyl amino acid transferase
OD	Optical Density
OPA	*o*-phthalaldehyde
OSB	*o*-succinylbenzoate
OSBS	*o*-succinylbenzoate synthase
Ox	Oxidated

PAGE	Polyacrylamide Gel Electrophoresis
Pank	Phosphate transferase
PAT	Phosphinothricin acetyl transferase
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PEP	Phospho Enol Pyruvate
Phg	Phenlyglycine
PLP	Pyridoxal 5'-Phosphate
*p*NP	*para*-nitro phenol
PNP	Purine nucleoside phosphorylase
*p*NPA	*para*-nitro phenol acetate
*p*NPB	*para*-nitro phenol butyrate
*p*NPP	*para*-nitro phenol palmitate
PPM	Phosphopentomutase
Prg	Propargyl-glycine
PT	Phosphinothricin
PTT	Phosphinothricin tripeptide
QM	Quadruple mutant
r.m.s.	Root Mean Square
r.t.	room temperature
Red	Reduced
RNA	Ribonucleic Acid
rpm	Revolution Per Minute
SDM	Site Directed Mutagenesis
SDS	Sodium-Dodecyl Sulphate
SEC	Size Exclusion Chromatography
SHCHC	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate
SOC	Super Optimal Culture
TA	Transaminase
TB	Terrific broth

TDA	Thermal Denaturation Assay
TEAA	Triethylammonium Acetate
TOF	Time Of Flight
TON	Turn Over Number
Tris	Tris(hydroxymethyl)aminomethane
UV-Vis	Ultraviolet Visible
V_{max}	Maximum Velocity
WT	Wild type
Y	Yield
α-KA	alpha-keto acid

Chemical Glossary

NCAAs and derivatives used in this experimental work. To identify the various compounds a classification system based on a letter plus number code was employed. The letter identifies the specific amino acids, based on the side chain; while the number indicates if the compound is free, or it has been derivatised, as reported in Figure G1.

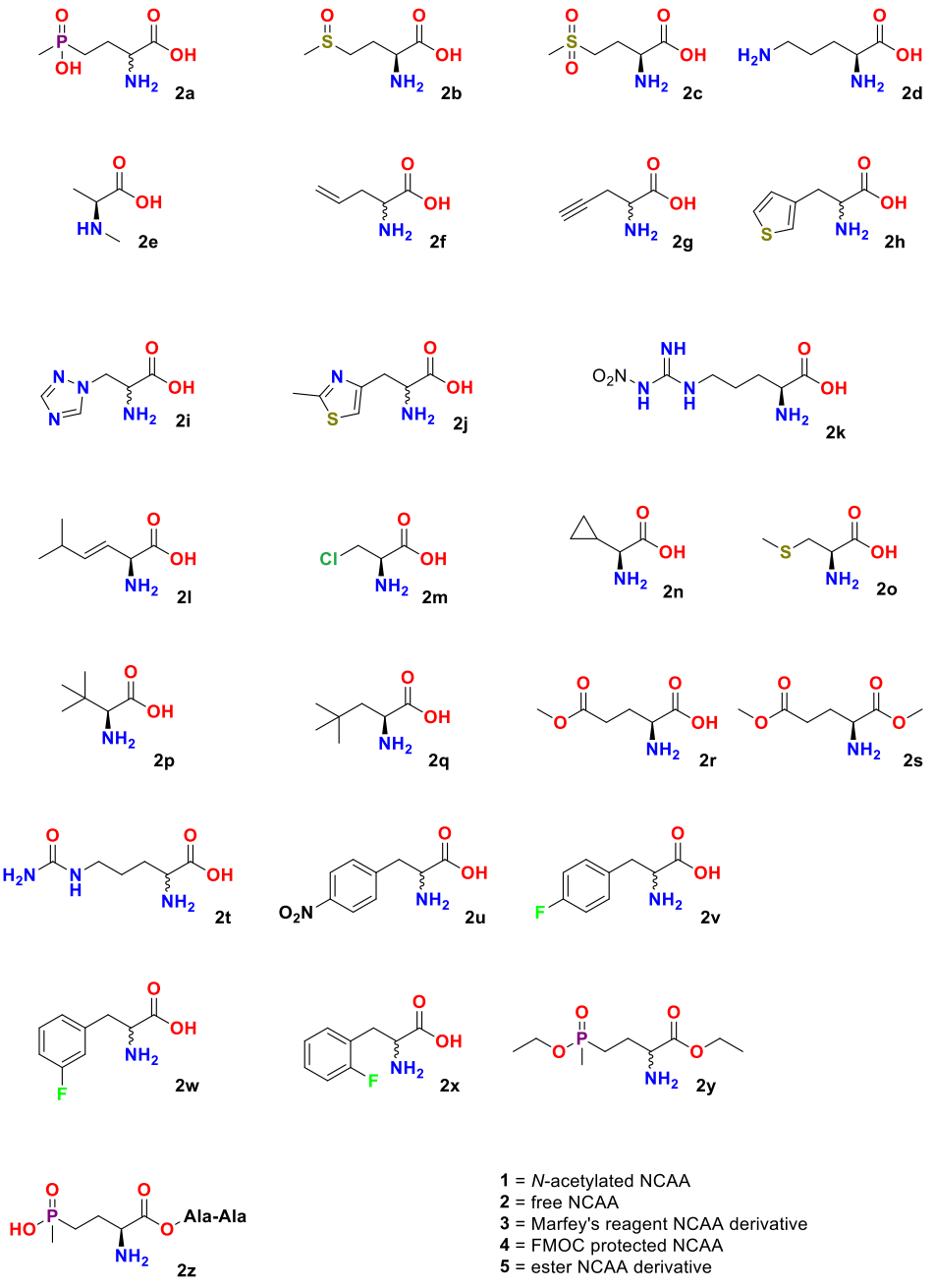


Figure G1 Classification system used in this experimental work to identify the various NCAAs derivatives employed. All NCAAs reported in their L-configuration.

Only two exception are the *para*-chloro and bromo phenylalanine derivatives, which are classified as reported in Figure G2.

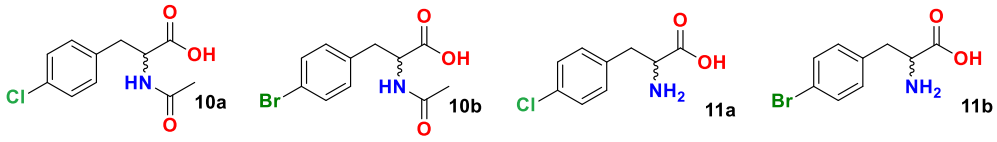


Figure G2 Classification system used in this experimental work to identify the *para*-chloro and *para*-bromo phenylalanine derivatives.

For the racemic compounds any change in the sample enantiopurity, following enzymatic resolution, will be spedified in the text.

Table of Contents

Lay Summary 1

Abstract 2

Dedication 4

Acknowledgments 6

Declaration 9

List of abbreviations 10

Chemical Glossary 15

Table of Contents 17

1 Introduction 23

1.1 Non-Canonical Amino Acids (NCAAs) 23

1.1.1 Applications 23

1.1.2 Classic strategies for NCAAs production 25

1.2 L-phosphinothricin (PT) 26

1.2.1 A peculiar natural product 26

1.2.2 Biosynthesis 26

1.2.3 Asymmetric synthesis of phosphinothricin 31

1.3 Biocatalysis 33

1.3.1 Overview 33

1.3.2 Industrial biocatalysis 34

1.3.3 Biocatalytic synthesis of amino acids 38

1.3.4 Future challenges 44

1.4 N-Acetyl-Amino Acid Racemase 45

1.4.1 Overview 45

1.4.2 The enolase superfamily 46

1.4.3 Role, structure and mechanism 48

1.4.4 Application 50

2 Aims 55
3 The *N*-acetyl-L-ornithine deacetylase ArgE 56
3.1 Aims 57
3.2 Expression and purification of un-tagged ArgE 57
3.3 Expression and purification of N-HisTag ArgE 59
3.3.1 High-throughput colorimetric L-amino acid oxidase assay 61
3.3.2 ArgE kinetic analysis 62
3.4 Kinetic resolution of *N*-acetyl methionine 64
3.5 Metal binding studies 66
3.5.1 Metal content of purified ArgE 66
3.5.2 Effect of Zinc and Cobalt on ArgE activity 67
3.6 ArgE Thermostability 69
3.6.1 Thermal Denaturation Assay 69
3.6.2 Correlation between reaction temperature and ArgE activity 70
3.7 ArgE activity towards *N*-Ac-PT 71
3.7.1 Synthesis and Crystal Structure 71
3.7.2 Kinetic Resolution of N-Ac-PT 72
3.7.3 Inhibition tests 72
3.8 Development of a quantitative assay for ArgE 75
3.8.1 Marfey's Reagent Derivatization 75
3.8.2 ^{1}H-NMR assay 78
3.9 Expanding ArgE substrate scope 83
3.9.1 Non canonical *N*-acetyl amino acid library synthesis 83
3.9.2 ArgE catalysed biotransformations 84
3.9.3 PT and Glu Studies 86
3.9.4 Kinetic Resolution on *N*-Ac-NCAAs 87
3.9.5 Determination of kinetic parameters of ArgE 90
3.10 Homology model and Mechanistic Studies 91
3.10.1 Homology Model 91
3.10.2 ArgE E144A, E145A and E169A Characterization 95

3.10.3 ArgE D112A, H195I and R264A Characterization 97
3.10.4 Proposed Mechanism 99
3.11 Crystal Trials 100
3.11.1 N-HisTag ArgE WT 101
3.11.2 C-HisTag ArgE WT 101
3.11.3 Non-Tagged ArgE WT 101
3.11.4 Crystal Optimization 103
3.12 Engineering of ArgE *via* random mutagenesis 105
3.12.1 *In vivo Arabidopsis thaliana* assay 106
3.12.2 OPA Assay 106
3.12.3 ArgE mutant library creation and screening 108
3.13 Conclusion and Future Work 110
4 The *N*-acetyl-bialaphos deacetylase (*Sv*Dea) 113
4.1 Aims 113
4.2 Expression and purification of N-His tagged *Sv*Dea. 114
4.3 Thermal denaturation assay (TDA) 115
4.4 Optimization of the purification of N-HisTag *Sv*Dea 116
4.5 Assaying *Sv*Dea activity and kinetic analysis 117
4.5.1 The *p*-NO_2 phenol esterase assay 117
4.5.2 The L-AAO/HRP coupled assay 120
4.6 Kinetic resolution of *N*-acetyl PT and derivatives 121
4.6.1 Substrate synthesis 121
4.6.2 Biotransformations 121
4.7 Structural studies. 123
4.7.1 Crystal Trials 123
4.7.2 Homology Model 124
4.8 Conclusion and Future Work 127
5 The *N*-Acetyl Amino Acid Racemase (NAAAR) 129
5.1 Aims 130
5.2 NAAAR WT and DM purification and characterization. 130

5.3 NAAAR WT and DM kinetic analysis 132
5.4 NAAAR DM Reaction Monitoring 134
5.5 Coupling of ArgE and NAAAR DM 135
5.5.1 DKR of *N*-Ac-DL-Met 135
5.5.2 Comparison between *E. coli* ArgE and *T. litoralis* L-Acylase 138
5.5.3 DKR of *N*-Ac-Non-Canonical-AAs 140
5.6 NAAAR QM expression, purification and characterization 142
5.7 Investigating the NAAAR QM substrate scope 144
5.7.1 Synthesis of N-Ac-AAs substrates 144
5.7.2 Small scale screening 144
5.8 DKR Scale-Up 148
5.9 Conclusions and Future Work 151
6 Materials and Methods 153
6.1 Materials and Reagents 153
6.1.1 Overview 153
6.1.2 Competent cells 153
6.1.3 Plasmids 153
6.1.4 Growth media, antibiotics and induction 154
6.1.5 Buffers 154
6.2 Molecular Biology 155
6.2.1 Site Directed Mutagenesis 155
6.2.2 ArgE Directed Evolution 156
6.2.3 *Sv*Dea Subcloning 158
6.2.4 ArgE Subcloning 159
6.2.5 Agarose gel electrophoresis and extraction 159
6.2.6 Digestion 160
6.2.7 Ligation 160
6.2.8 Heat shock transformation 160
6.2.9 DNA miniprep 161
6.2.10 Analytical digest and sequencing 161

6.3 Protein production 161
6.3.1 Expression test protocol 161
6.3.2 ArgE mutant library screening 162
6.3.3 NAAAR variants expression[4] 162
6.3.4 ArgE Expression 162
6.3.5 *Sv*Dea Expression 163
6.3.6 TEV Expression 163
6.4 Protein purification 163
6.4.1 Cell free extract preparation 164
6.4.2 Anion exchange 164
6.4.3 Nickel NTA purification 164
6.4.4 TEV cleavage 165
6.4.5 Size exclusion chromatography 165
6.5 Protein characterization 165
6.5.1 Mass spectrometry 165
6.5.2 SDS-PAGE electrophoresis 166
6.5.3 Protein concentration 166
6.5.4 ICP-MS 166
6.6 Assays protocols 168
6.6.1 The L-AAO assay 168
6.6.2 The *p*-NO_2 phenol assay[6] 168
6.6.3 The OPA assay 168
6.6.4 Marfey's derivatization protocol[8] 169
6.7 Biotransformations 169
6.7.1 ArgE biotransformations protocol 169
6.7.2 *Sv*Dea hydrolysis of *N*-Ac-L-amino acids 169
6.7.3 NAAAR/deacetylase DKR 170
6.7.4 Gram scale enzymatic reaction 170
6.8 Organic Synthesis 170
6.8.1 Deprotection of FMOC amino acids 170

6.8.2 Ester hydrolysis 171
6.8.3 Amino acid acetylation 171
6.8.4 Marfey's reagent synthesis[8] 172
6.9 Compound Characterization 172
6.9.1 NMR spectroscopy 172
6.9.2 LC-ESI MS spectrometry 173
6.9.3 Chiral HPLC 173
6.10 Crystal trials 173
7 Conclusions 175
8 References 178
9 Appendices 187

1 Introduction

1.1 Non-Canonical Amino Acids (NCAAs)

1.1.1 Applications

Given their inherent chirality and structural variety, enantiopure canonical and non-canonical amino acids (NCAAs) have always played a fundamental role as synthons for the production of fine chemicals and additives.[1] Since NCAAs are not genetically encoded, they cannot be naturally incorporated into proteins during translation. For this reason, NCAAs are often referred as non-proteinogenic or unnatural.[2] However, the term non-canonical is more generic, because it comprises both natural product and synthetic AAs. Indeed, the vast majority of known NCAAs are produced by animals, plants and microorganisms as secondary metabolites and employed as intermediates in various biosynthetic pathways (e.g. citrulline and ornithine are formed during the urea cycle), [3] or as agents for pests' control.[4,5]

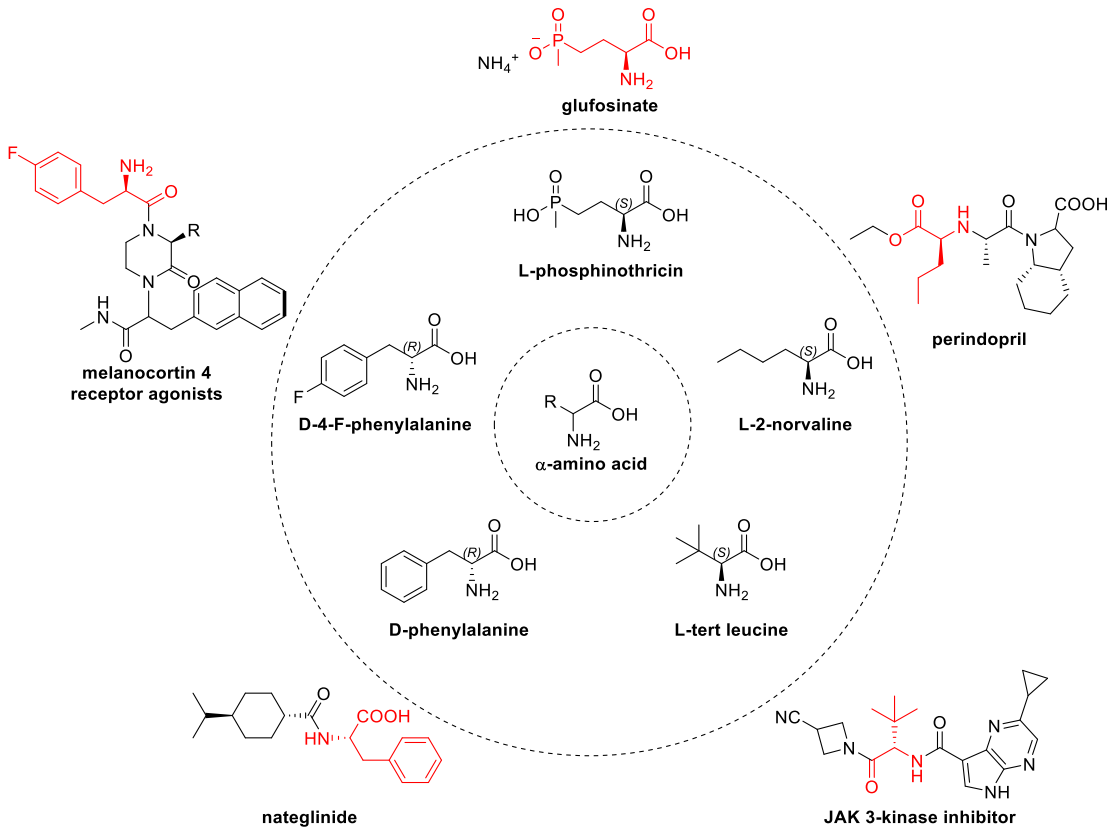


Figure 1.1 Example of NCAAs employed in the preparation of pharmaceuticals and agrochemicals.[6,7] Figure adapted from Zhou *et al.*[8]

NCAAs may possess a peculiar group on their side chain, a D-stereoconfiguration, or the chiral centre moved further along the carbon chain. Thanks to their structural variety, NCAAs have found numerous applications in various fields.[9] As biologically active enantiopure compounds they have been employed as chiral building blocks for the synthesis of pharmaceuticals and agrochemicals (Figure 1.1).[6] As of 2010,[10] 12% of the top 200 drugs on the market contained a non-proteinogenic amino acid.

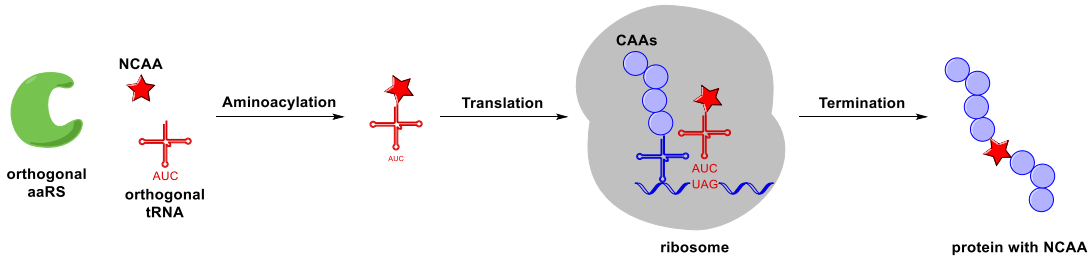
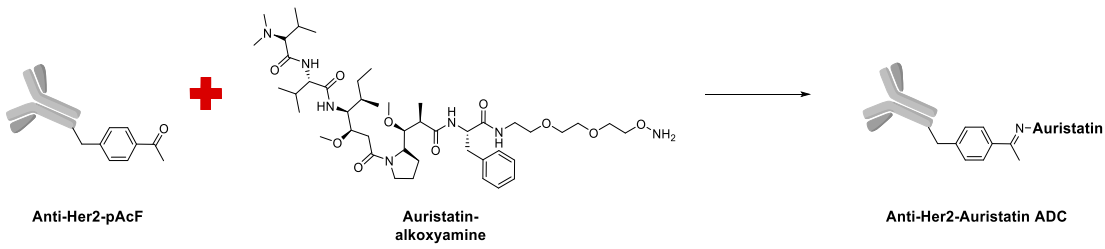


Figure 1.2 Genetic code expansion strategy for the incorporation of a NCAA into a protein sequence. An orthogonal aminoacyl-tRNA synthetase (aaRS) catalyse the coupling of the NCAA with the orthogonal tRNA molecule consuming ATP. The loaded tRNA is then bound to the ribosome during translation to incorporate the NCAA in the protein sequence at the desired position. Adapted from Zhao, Burke and Green.[11]

The global substitution of a canonical amino acid with a non-canonical analogue is an important tool used to tune the chemical and physical properties of proteins and enzymes. For instance, the substitution of methionine with selenomethionine[12] allows heavy atom incorporation and the resolution of a protein X-ray structure; a necessary strategy whenever molecular replacement is not a viable option. Recently NCAAs have also been employed as tools for protein engineering, thanks to the development of various genetic code expansion strategies, which allow to selectively introduce a NCAA at the desired position in a protein sequence (Figure 1.2).[2,11,12]



Scheme 1.1 Synthesis of Anti-Her2-Auristatin. The Anti-Her2 antibody was modified to incorporate the non-canonical amino acid *p*-acetyl phenylalanine (pAcF). The acetate group of the NCAA reacts with the auristatin-alkoxyamine drug *via* electrophilic addition and oxime formation to produce the ADC.[13]

These techniques can be employed to design enzymes with enhanced activity,[14] as recently reported by Green,[15] non-natural cofactors (e.g. artificial metalloenzymes) or novel mechanisms.[16–18] Genetic code expansions are also employed in the development of biotherapeutics with NCAAs, such as antibody-drug conjugates (ADC), therapeutic peptides and protein-based vaccines, with increased stability, potency and selectivity (Scheme 1.1).[13]

1.1.2 Classic strategies for NCAAs production

The increase of the applications of NCAAs in recent years led to a rise in the market demand of these compounds. The majority of canonical L-amino acids are prepared at large scale *via* fermentation processes using genetically modified bacterial strains including *Escherichia coli* and *Corynebacterium glutamicum*.[19] This technique cannot be applied for the production of enantiopure NCAAs, since their biosynthetic pathways are often complex and not well studied, which makes metabolic engineering quite challenging.[3] However, a few exceptions have been described in the literature, such as for β-phenylalanine production. [20]

Following the publication of the first amino acid synthesis, reported by Strecker in 1850,[21] a wide range of strategies have been developed for the preparation of these compounds. According to the work of Nàjera and Sansano[22] the stereoselective synthesis of α-amino acids can be achieved on a small scale, with good to excellent results, using four main strategies (Figure 1.3). However, most of the enantioselective catalysts employed in these reactions are either metal complexes with chiral ligands or optically pure organocatalysts and their application at industrial scale is limited by the high-cost of production, the toxicity of the transition metals and the complex synthesis of the ligands.

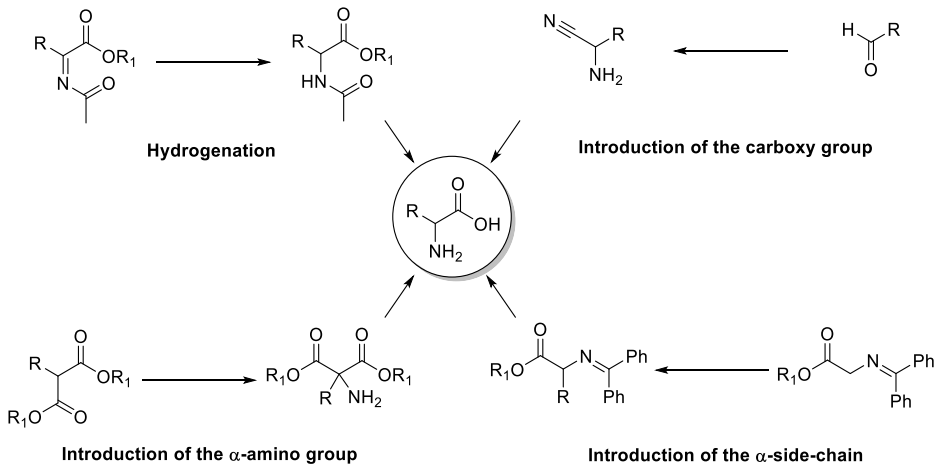


Figure 1.3 The four main strategies used for the chemocatalytic synthesis of amino acids.[22]

Currently the majority of non-proteinogenic amino acids are chemically synthesised and sold as racemic mixtures.

1.2 L-phosphinothricin (PT)

1.2.1 A peculiar natural product

L-phosphinothricin (PT) is a secondary metabolite produced by various species of actinomycetes and a NCAA structural analogue of glutamic acid (Figure 1.4). Being the only phosphinic acid-derived natural product isolated, PT belongs to the family of the bioactive C-P compounds. Phosphinic acids have a structure similar to the phosphate anhydrides, however the O-P-O bond is substituted by a stronger C-P-C bond, which is able to resist harsh thermal and chemical treatments.[23] The bioactivity of C-P natural products is given by their structural similarities with carboxylic acids and anhydrides; these compounds usually acts as strong inhibitors competing with the substrate for the enzyme active site.[23]

1.2.2 Biosynthesis

In the L-PT biosynthetic pathway the free amino acid is never directly produced, since L-PT is always incorporated in the structure of tri- or tetra-peptides (Figure 1.4). Examples include bialaphos (PT-Ala-Ala or PTT) isolated from *Streptomyces viridochromogenes*, phosalacine (PT-Ala-Leu) by *Kitasatospora phosalacina* and trialaphos from *Streptomyces hygroscopicus* (PT-Ala-Ala-Ala).[23] These bioactive peptides use a so-called "Trojan Horse" mode of action, since they can be easily assimilated through the cell membrane of the target organism and hydrolysed by various endogenous non-selective peptidase, releasing the bioactive L-PT.

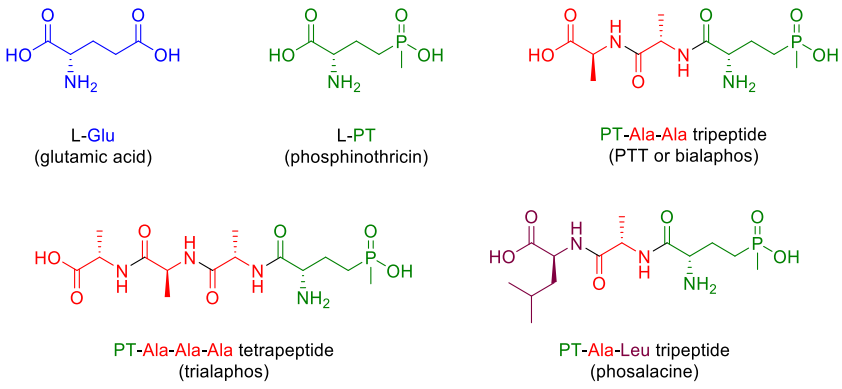


Figure 1.4 Structure of the L-PT containing non-ribosomal peptides isolated from *S. viridochromogenes* (bialaphos, PTT), *S. hygroscopicus* (trialaphos) and *K. phosalacina* (phosalacine).

These natural products belong to the family of non-ribosomal peptides (NRP),[24] since their synthesis is completely independent from the mRNA and ribosomes, but is under the control of NRP-synthetases (NRPS). This large multi-enzymatic system formed by various units, which are dedicated to the recognition and activation of amino acid residues and peptide bond formation. Isolation of the PTT gene cluster from *S. viridochromogenes* Tü494,[25,26] (Figure 1.5 and Table 1.1) allowed for the elucidation of the steps involved in the tripeptide biosynthesis from phosphoenolpyruvate (PEP, Scheme 1.2). Two molecules of PEP serve as starting material for the formation of *N*-acetyl demethyl-phosphinothricin (NAcDMPT); twelve enzymes are directly involved in the synthesis of this L-PT precursor. NAcDMPT is then activated *via* adenylation (adenylation A module) by the first module of the NRPS (PhsA), while the other two larger modules (PhsB and PhsC) activate, *via* thioester formation (PCP), and add an alanine residue (condensation C module) the growing peptide (Scheme 1.3).[24,27] The formation of the peptide bonds is followed by the methylation of the phosphinic acid and the hydrolysis of the acetate group, yielding bialaphos as final product (Scheme 12).

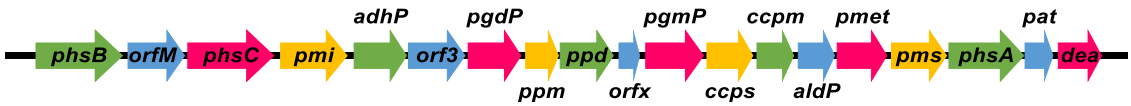
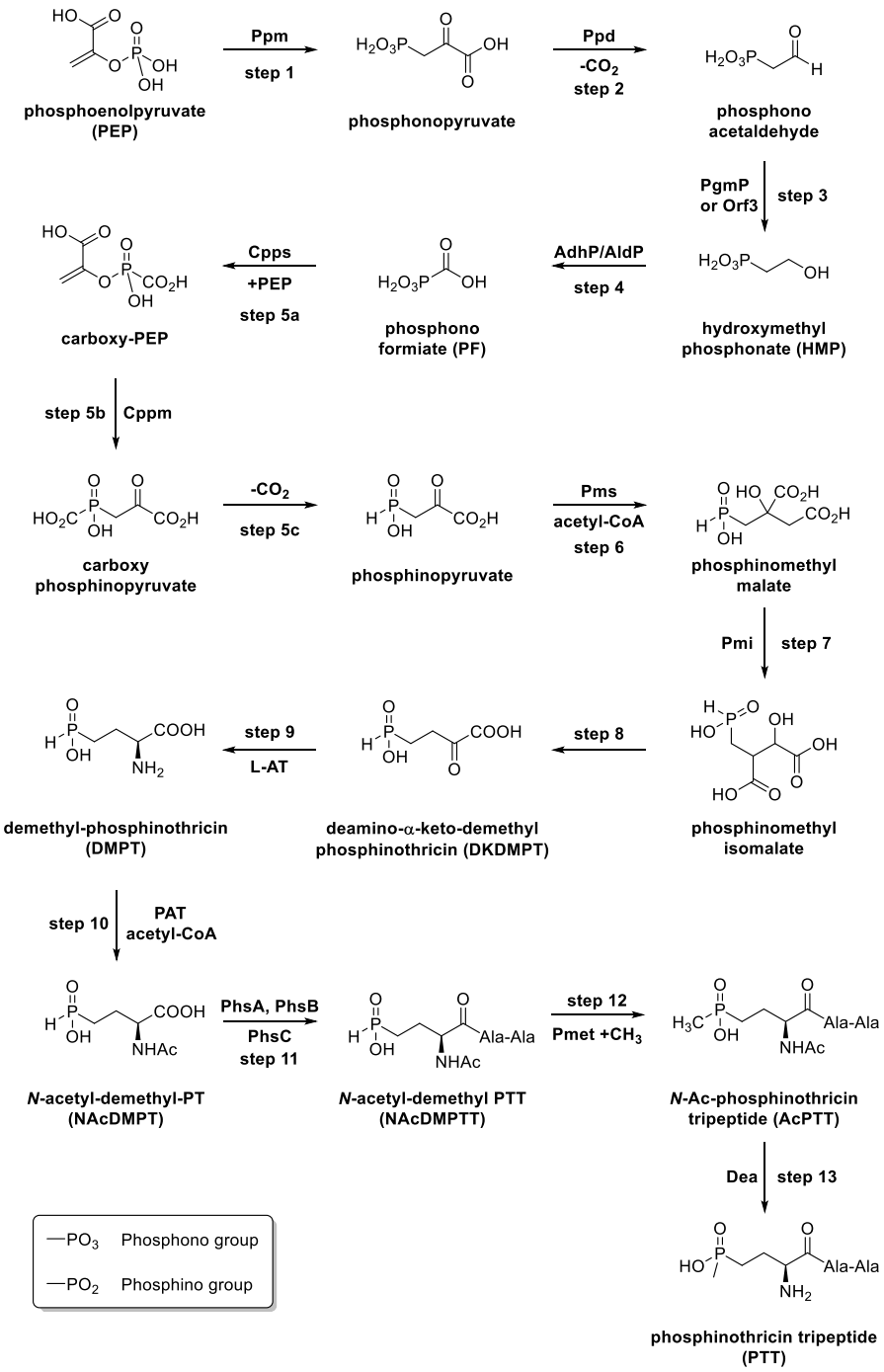


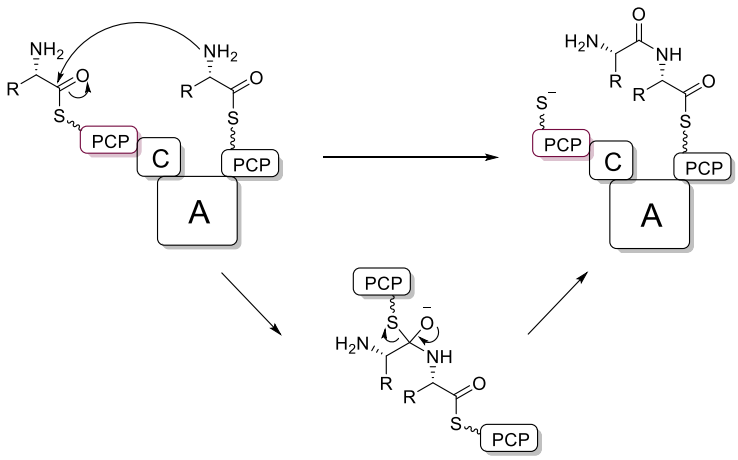
Figure 1.5 The PTT biosynthetic gene cluster from *S. viridochromogenes* Tü494 with the ORFs drawn in scale, according to Schwartz *et. al*. [26]

Abbreviation	Enzyme
Ppm	phosphoenolpyruvate mutase
Ppd	phosphonopyruvate decarboxylase
PgmP	putative phosphoglycerate mutase
AdhP	putative alcohol dehydrogenase
AldP	putative aldehyde dehydrogenase
Cppm	carboxy-PEP phosphonomutase
Cpps	carboxy-PEP synthase
Pms	phosphinomethylmalic acid synthase
Pmi	phosphinomethyl malate isomerase
L-AT	putative aminotransferase
PAT	phosphinothricin *N*-acetyltransferase
PhsA	phosphinothricin tripeptide synthetase A
PhsB	phosphinothricin tripeptide synthetase B
PhsC	phosphinothricin tripeptide synthetase C
Pmet	P-methylase
Dea	deacetylase

Table 1.1 List of the sixteen enzymes, abbreviations and full names, involved in the biosynthesis of PTT in *S. viridochromogenes* Tü494.



Scheme 1.2 Proposed *S. viridochromogenes* Tü494 thirteen steps biosynthetic pathway for the synthesis of PTT.[26]



Scheme 1.3 Mechanism for the formation of the peptide bond catalysed by the condensation domain C of NRPs. Following amine attack at the carbonyle, the process passes through the formation ofa tetrahedral intermediate, which collapses forming the desired dipeptide product.

It was found that L-PT displays antibacterial, fungicidal and herbicidal activity, being a strong inhibitor of the enzyme glutamine synthetase (GS).[28] Racemic PT is employed as the bioactive ingredient of various commercial herbicide formulations (e.g. Basta® and Finale® produced by Bayer)[29] and used for weed control in both uncultivated and crop fields. It is sold as a soluble concentrate (200 g/L) at 8.80 $/L and applied on the crops *via* spraying the diluted formulation over the fields. L-PT can also cause acute toxicity in humans *via* ingestion; a dose of ~300 mg/Kg of Basta® produces severe damage to the nervous and circulatory systems and it can be fatal if not treated adequately .[28] However, its efficacy as an antibiotic is quite poor, since the host tissues can easily provide glutamine to the invading organism, effectively nullifying L-PT toxicity.

The target GS enzyme plays a fundamental role in the metabolism of plants and other organisms, since it is able to fix ammonia, using glutamic acid to synthesize the amino acid glutamine.[30] The GS enzymes have been divided in two groups; the type I found in prokaryotes is a dodecamer composed by twelve identical units, while the type II isolated in eukaryotes (animals and plants) is a decamer (Figure 1.6).[30,31] Despite the structural differences, the two GS groups share a common two-step mechanism (Scheme 1.3) in which the glutamic acid bound to the enzyme is first activated by transfer of a phosphate group from ATP to the γ-carboxylate of the substrate; an ammonium ion is then deprotonated by the D56 residue situated on the adjacent chain of GS. The ammonia generated undergoes a

nucleophilic substitution with the activated substrate to form glutamine with the release of a phosphate group.

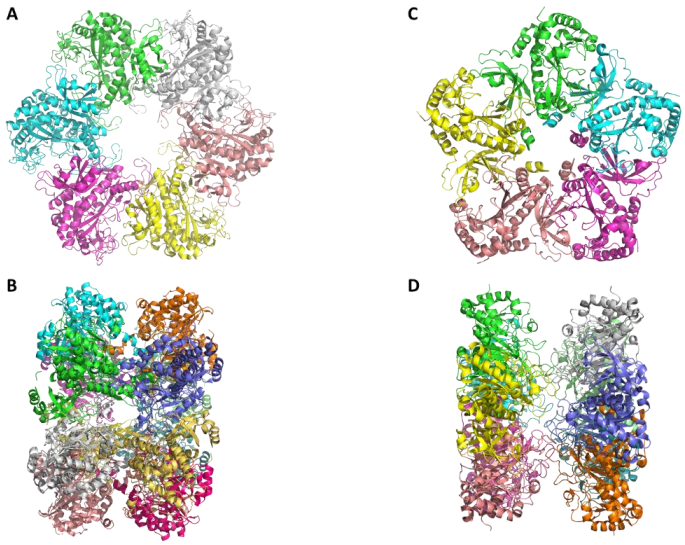
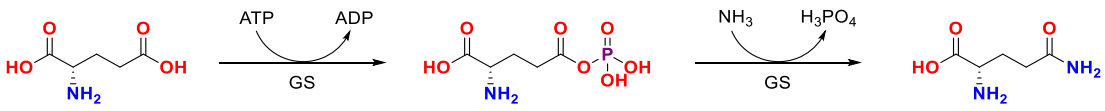


Figure 1.6 Bacterial and Plant GSs resolved crystal structures. **A)** Crystal structure of *Mycobacterium tuberculosis* GS (1HTQ), 6 identical chains interact to form a hexagonal agglomerate. **B)** Two agglomerates stack together to give the dodecameric quaternary structure of the enzyme. **C)** Crystal structure of Maize GS (2D3C), 5 identical units interact to form a pentagonal agglomerate. **D)** Two agglomerates stack together to give the decameric quaternary structure of GS.

L-PT can compete with L-glutamate for the GS active site; upon PT binding, the amino acid phosphinic group is phosphorylated by the enzyme using ATP, generating a stable intermediate which is not able to undergo nucleophilic substitution by ammonia, effectively inhibiting the enzyme activity (Figure 1.7). This cause an increase in the concentration of ammonium ions inside the cell, leading to its death.[30]



Scheme 1.4 Two-step mechanism of the enzyme GS. Following the phosphorylation of the substrates γ-carboxylate, a nucleophilic ammonia attacks the activated intermediate to produce glutamine *via* a nucleophilic acyl substitution.

Phosphinothricin is able to bind to both eukaryotic and prokaryotic GS enzymes; however various PT resistant bacterial strains have been isolated during the years (*Rhodococcus, Alcaligenes faecalis, Streptomyces* and *E. coli*). These bacteria have evolved various strategies for PT detoxification; when the NCAA enters the organism, it is immediately converted to an inactive derivative, usually *via* oxidative deamination, catalysed by an amino acid oxidase

(AAO) or transamination. Another strategy used by these microorganisms is the *N*-acetylation of L-PT catalysed by a phosphinothricin acetyl transferase (PAT) enzyme. [32–34] The PAT isolated from *S. viridochromogenes* and *S. hygroscopicus* are well characterised and have been employed to create various PT resistant crops used in agriculture.[35]

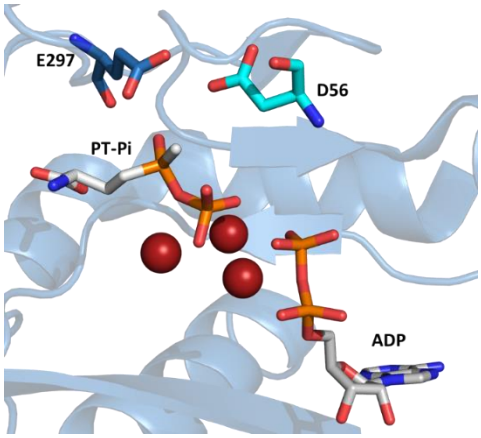
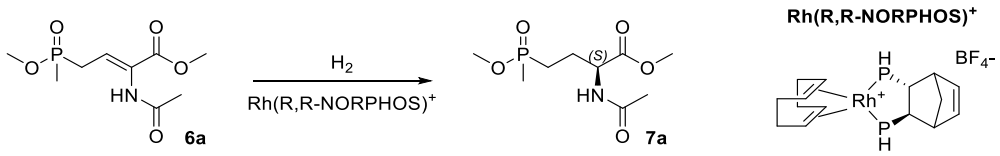


Figure 1.7 Crystal Structure of the Maize GS (2D3C) in complex with three Mn^{2+} ions (red spheres), **ADP** and phosphinothricin phosphate (**PT-Pi**). The GS active site is formed by residues located on two adjacent chains; **D56** (cyan) from chain B is involved in the ammonium ion binding and deprotonation, while **E297** (blue) from chain A plays a role in substrate recognition.

Commercials herbicide are classified accordingly to which part of the plant metabolism they are going to disrupt, usually *via* inhibition of a target enzyme. Common mode of actions includes photobleaching and the disruption of cell division, carotenoid biosynthesis, photosynthesis, or lipid synthesis (acetyl-CoA carboxylase inhibitors).[36] Other herbicides with a similar mode of action (distruption of amino acid synthesis) to L-PT are the inhibitors of 5-enolpyruvyl-shikimate-3-phosphate synthase (EPSP synthase), involved in the biosynthesis of aromatic amino acids (Phe, Tyr and Trp) and of acetolactatesynthase (ALS synthase), which intervenes in the production of branched AAs (Val, Leu, Ile).[37]

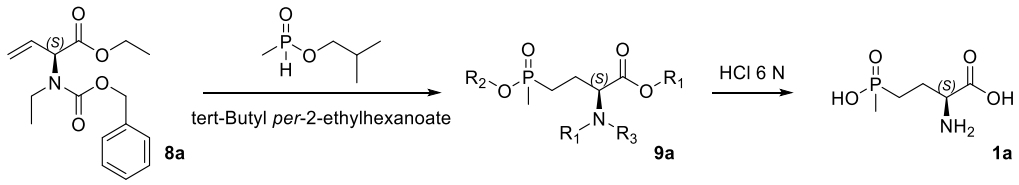
1.2.3 Asymmetric synthesis of phosphinothricin

As with many other natural products, the two stereoisomers of PT possess different proprieties; while the L-enantiomer is a potent inhibitor, the D-isomer is completely inactive. For this reason, the development of an efficient enantioselective route for the production of L-PT has been an important focus of research. Given the complexity of the phosphinothricin biosynthetic pathway metabolic engineering was not a viable option; however, two synthetic strategies have been reported by Zeiss in the 1990s.



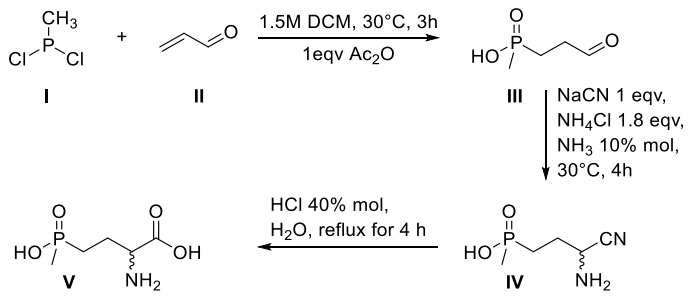
Scheme 1.5 Enantioselective hydrogenation of PT prochiral intermediate **6a**, catalysed by optically active Rh(I) bis-phosphines complexes. The ligand *R,R*-NORPHOS provided a 90.8% *ee* in the synthesis of *S*-**7a** in MeOH/H_2O with a 150:1 substrate/catalyst ratio loadings.[38]

The first stereoselective route[38] consisted of a six-step synthesis (Scheme 1.4) where the prochiral **6a** intermediate undergoes a hydrogenation catalysed by chiral rhodium-phosphine complexes; the reduced product is then hydrolysed to give optically pure PT ammonium. By changing the ligand on the chiral complexes, it was possible to selectively produce either the *S* or *R* enantiomer with up to 91% *ee*.



Scheme 1.6 Two-step synthesis of L-PT from L-vinyl glycine derivative **8a**.[39]

A second three-step synthesis of L-PT using L-vinyl glycine derivatives as starting material[39] was reported in 1992 (Scheme 1.5). The desired product was obtained in moderate to good yield (57-97%) and high enantiopurity (92-97% *ee*); unfortunately, the preparation of vinyl glycine derivatives from L-methionine, or L-glutamic acid, requires the use of toxic reagents, such as lead acetate, and yields small amounts of the desired product.



Scheme 1.7 Three-step synthesis of racemic PT from methyl-di-chloro phosphate and acrolein, *via* Strecker reaction.[40]

Both stereoselective strategies rely on the use of toxic reagents or expensive catalysts and could not be applied for industrial scale production. Currently PT is synthesised using a classic Strecker reaction (Scheme 1.6)[40] to introduce the α-amino and α-carboxylate groups onto

the carbon chain and the final product is sold as a racemic mixture. For every application of L-PT based herbicide, 50% of the product used is wasted. Studies of various cultivations revealed that ~90% the of D-PT applied[41] is not metabolised by plants, but an accumulation of this stereoisomer can potentially be dangerous for the environment. An alternative solution to the classic organic synthesis strategy is required to develop an industrial scale process for L-PT production.

1.3 Biocatalysis

1.3.1 Overview

Biocatalysis, or the use of enzymes to catalyse chemical reactions, has been employed for centuries for food and alcoholic drink production; even if its true potential has been discovered more than a century ago, only recently biocatalysis has been applied in the development of novel synthetic strategies, especially for the design of chiral compounds (Figure 1.8). In the last thirty years the progress achieved in the field of molecular biology and gene technology has made it possible to isolate, express and evolve the target enzyme to obtain the desired process-tailored biocatalysts. Bornscheuer[42,43] recently announced the rapid approach of a fourth wave of biocatalysis, which would focus on the expansion of the pool of biocatalysed reactions and on the development of efficient enzymatic cascade *via* metabolic engineering to achieve a cost-effective strategy for the synthesis and isolation of challenging natural products targets.

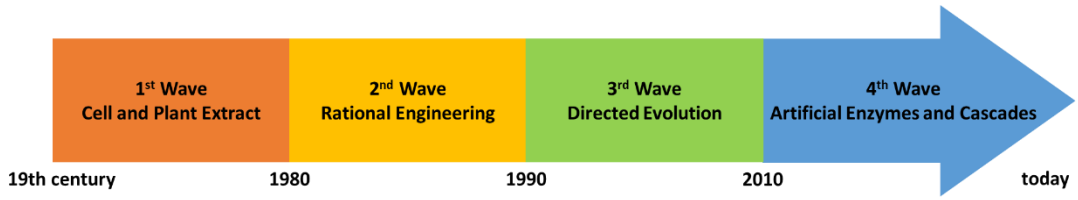


Figure 1.8 The four waves of biocatalysis. Bornscheuer proposed this classification to describe the progress achieved in the field and the current state of the art. [42,43]

While the preparation of optically pure molecules still represents a challenge for organic chemistry, enzymes are characterised by intrinsic high chemo, regio and enantioselectivity. Furthermore, the set-up for a typical biocatalytic reaction is quite straightforward and "safe" since enzymes work at low temperatures (20-37°C), in water-based buffer solutions, at neutral pH (6-8) and atmospheric pressure. [44] The mild reaction conditions, coupled with the use of biodegradable proteins as catalysts, classifies biocatalysed processes as

environmentally friendly; a factor which contributed to the increase in popularity of this filed, given the growing concern for the current state of the environment and the need to eliminate the use of hazardous reagents in manufacturing.[45]

However, the application of biocatalysis at industrial scale has been slowed by several limitations of enzymatic processes, such as the low stability of proteins operating outside, far from physiological conditions, the limited range of enzyme-catalysed reactions and the low tolerance for organic solvents, which make the synthesis and recovery of hydrophobic compounds quite challenging. Furthermore, the high selectivity of biocatalysts is also their Achilles heel, since they are usually characterised by a narrow substrate scope.[44]

Thanks to the pioneering work of Stemmer,[46,47] Reetz[48,49] and Arnold[50,51] (Nobel prize for chemistry, 2018) a wide range of novel techniques have been developed and successfully employed to engineer enzymes *via* directed evolution, or rational design. Directed evolution is best described as an *in vitro* Darwinian selection strategy, where a range of techniques could be employed to generate a large number of variants of the target enzyme with mutations at random positions in the amino acid sequence; this mutant library is then screened to isolate the best variant with the desired properties (i.e. activity, stability, selectivity). While the rational design approach relies on computational analysis and protein structural information, to identify the target amino acid residues to mutate.

Both strategies have been successfully employed not only for the improvement of enzyme stability and activity, but also to alter the substrate specificity and to gain access to novel biocatalysed reactions, including Diels-Alder condensations, Kemp elimination and cyclopropanation, with few to no previous examples discovered in nature.[52]

1.3.2 Industrial biocatalysis

Progress in the field of molecular biology combined with the successful application of enzyme engineering and immobilization has allowed researchers to overcome some of the classic limitations of biocatalysis, as a result biocatalysed processes became more suitable for industrial applications.

Currently, enzymes are employed in the manufacture of food (prebiotics and low-calorie sweeteners productions), cosmetics (emollient ester synthesis), textile (dyeing and bleaching), pulp and paper (lignin and hemicellulose removal) and fine and bulk chemicals (acrylamide and glycolic acid production).[53,54] In the last five year (2014-2019) ~200 patents,

which cover the use of all main enzyme classes, have been approved (Figure 1.9) and a growth in the enzyme market from 7$ to 10$ billion is expected by 2024.[55]

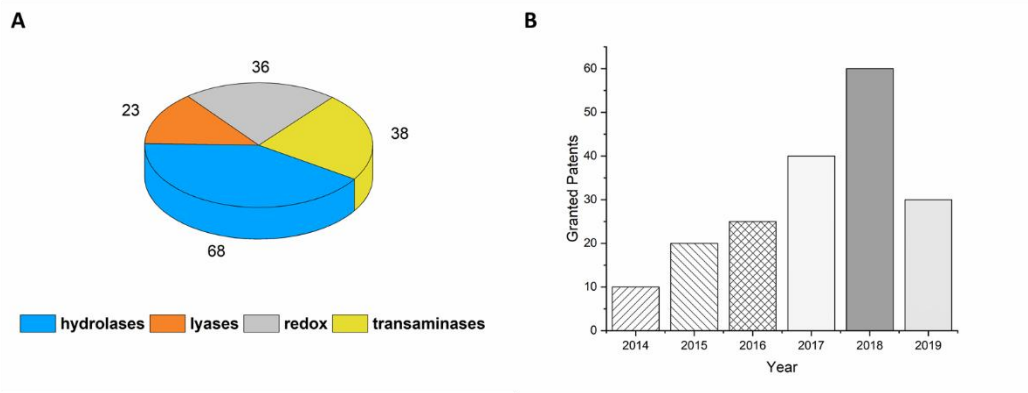
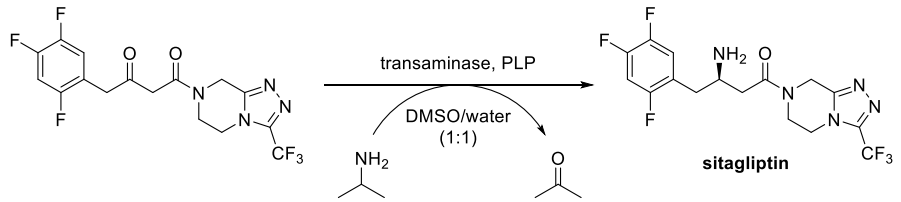


Figure 1.9 Distribution of patented enzymes. **A)** Number of approved patents for each major class of enzymes between 2014 and 2019. **B)** Distribution of granted patents, which cover the use of an enzymatic processes, over the last five years.[55]

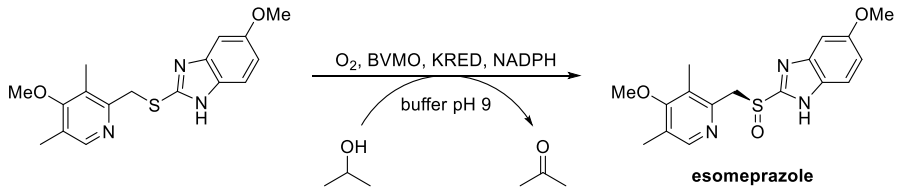
However, biocatalysis has mainly been exploited in the last decade in the pharmaceutical field, as a potent tool for the production of small stereoisomeric active pharmaceutical ingredients (API) *via* resolution (maximum yield 50%), or stereoselective introduction of a stereocenter (100% theoretical yield).[56,57] The exquisite regio- and enantio-selectivity of enzymes, the ability to tune the catalysts to match the desired properties (activity, stability), the recent development in bioinformatics (*de novo* design and biocatalytic retrosynthesis) and the decreasing cost of synthetic genes have contributed to the increasing interest that has caused many global pharmaceutical companies to take to this field. [58]

In 2010 Merck and Codexis were awarded the Green Chemistry Prize award for the biocatalytic production of sitagliptin (treatment of diabetes mellitus type II) *via* transamination (Scheme 1.7).[59] The transaminase (TA) ATA-117 was engineered using rational design in combination with saturation mutagenesis to isolate an active mutant towards the prochiral precursor: the prositagliptin ketone. The final variant contained 27 mutations (*R*-TA-27 mut), which were necessary not only to modify the TA substrate scope and activity, but also to improve its stability towards the required reaction conditions: high concentration of the amino donor isopropylamine (IPA) and DMSO. The final biocatalytic process provides sitagliptin with an increase in the final yield and *ee*, along with a reduction of waste and cost production, compared to the classic chemocatalytic strategy, where an asymmetric Rh catalysed hydrogenation was necessary to enantioselectively convert the ketone into the desired product.



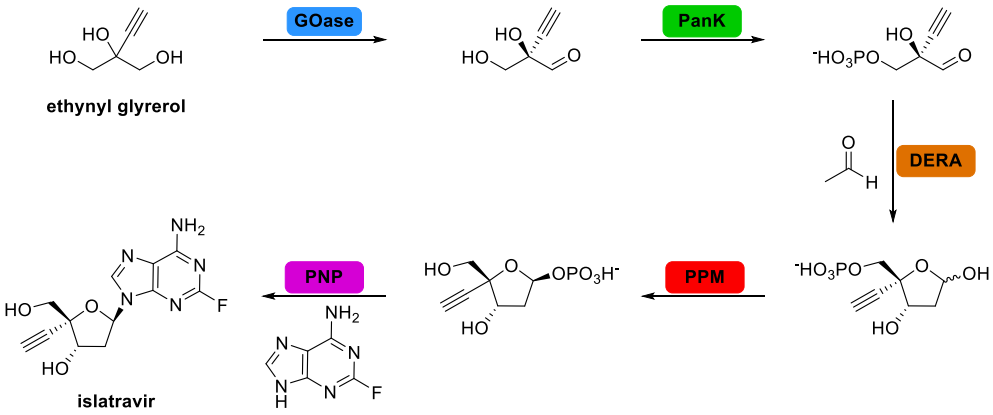
Scheme 1.8 Asymmetric synthesis of sitagliptin from the prochiral ketone precursor *via* transamination. [59]

The biocatalytic synthesis of sitaglipin has become a milestone for the employment of enzymes in API manufacture, since 2010 numerous applications of engineered enzyme in drug discovery have been reported in literature.[56] In 2015 Codexis reported the used of an evolved Bayer-Villiger monooxygenase (BVMO) for the synthesis of esomeprazole (narcolepsy treatment).[60] Esomeprazole belongs to a class of APIs containing a chiral sulfoxide group, which is usually obtained from the oxidation of a sulphur precursor. Unfortunately, classic chemosynthetic strategy relies on the use of the Kagan-Sharpless catalyst, which is plagued by various drawbacks: poor yield, low enantioselectivity and excessive oxidation, with the formation of the sulfone side-product. A BVMO enzyme with activity for the esomeprazole precursor, pyrmetazole, was isolated and subjected to various rounds of directed evolution. The final variant contained 41 mutations and displayed an 140000-fold increase in productivity and < 0.1% sulfone formation. The engineered BVMO was coupled with a ketoreductase (KRED) active towards isopropylalcohol for continuous cofactor (NADPH) regeneration to create an efficient cost-effective industrial scale process (Scheme 1.8).



Scheme 1.9 Engineered Codexis BVMO, coupled with KRED, for the catalysed oxidation of a sulphur to sulfoxide for the enantioselective synthesis of esomeprazole. [60]

Recently a biocatalytic retrosynthesis approach has been employed by Merck to create an efficient *in vitro* biocatalytic cascade, consisting of five engineered (Table 1.2) and four supporting enzymes for cofactor recycling, for the production of the anti-viral nucleoside analogue islatravir, developed for HIV treatment (Scheme 1.9).[61]



Scheme 1.10 Enzymatic cascade for the biocatalytic preparation of the anti-viral HIV drug islatravir from prochiral ethynyl glycerol. [61] GOase, *Fusarium graminearum* oxidase; PanK, *E. coli* phosphate transferase; DERA, *Shewanella halifaxensis* deoxyribose 5-phosphate aldolase; PPM, *E. coli* phosphopentomutase and PNP, *E. coli* purine nucleoside phosphorylase.

Biocatalytic cascades exploit the ability of various enzymes to work in cooperative conditions and are considered a promising strategy for the synthesis of small molecules with various functional groups since no intermediate isolation and protection/deprotection steps are required. Furthermore, enzyme coupling is an effective strategy to overcome thermodynamically unfavourable steps pushing the equilibrium towards product formation. According to the retrosynthetic analysis, five steps are required to obtain islatlavir starting from the cheap prochiral ethynyl glycerol. The *Fusarium graminearum* oxidase (GOase) and the *E. coli* phosphate transferase (PanK) catalyse respectively the oxidation and phosphorylation of an alcohol group of ethynyl glycerol for the desymmetrisation of the starting material. This is followed by an aldol consensation catalysed by *Shewanella halifaxensis* deoxyribose 5-phosphate aldolase (DERA) for the formation of the sugar moiety, which is subsequently phosphorylated by the *E. coli* phosphopentomutase (PPM). Finally, the *E. coli* purine nucleoside phosphorylase (PNP) catalyse the condensation between the sugar moiety and the synthetic nitrogenous base to yield islatravir. A combination of rational design and directed evolution was employed to evolve the selected enzymes and once the improved variants had been isolated, a strategy on how to best couple the mutants was developed. A total of nine enzymes: the five engineered biocatalysts and four auxiliary enzymes, required for efficient cofactor recycling and to push the equilibrium towards product formation, were employed for the one-pot biocatalytic production of islatravir with an overall yield of 51% and 99.5% diastereomeric ratio.

Enzyme	Evolution Focus	Rounds of Evolution	N° Residues Changed
GOase	stereoselectivity	12	34
Pank	activity	3	10
DERA	acetaldehyde tolerance	2	11
PPM	activity	2	5
PNP	activity	4	7

Table 1.2 Evolution of the five enzymes directly involved in islatravir synthesis.[61]

1.3.3 Biocatalytic synthesis of amino acids

Biocatalysis established itself as a competitive and effective strategy for the production of small molecules with a single stereocenter and polar groups (alcohols and amines). Given the ability of enzymes to tolerate a wide range of functional groups, without the need of the numerous protection and de-protection steps, typical of a chemosynthetic process. Indeed, biocatalysis has become a popular tool, even among chemists, for the production of small polar compounds, such as amino acids, which can be achieved with two main strategies: (I) asymmetric synthesis, or (II) kinetic resolution. A large variety of enzymes have been isolated and applied for the enantioselective synthesis of these important chiral building blocks.[62]

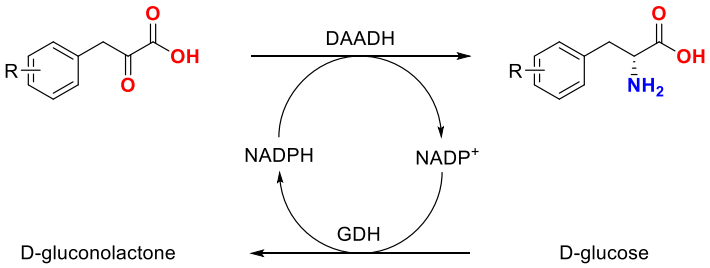
1.3.3.1 Asymmetric biocatalytic synthesis of amino acids

This strategy involves the stereoselective introduction of the primary amine group on a prochiral starting material, such as keto and unsaturated acids; this type of synthesis can be achieved by different classes of enzymes, each with a unique mechanism.

Amino Acid Dehydrogenases (AADH) catalyse the reversible reductive amination of α-keto acids (KA) to α-amino acids. The reduction is achieved using ammonia as amino donor and NAD(P)H as cofactor, which is converted to $NAD(P)^+$ during the process. Being quite expensive, a good cofactor recycling strategy needs to be employed to obtain an efficient and cost-effective synthesis on a large scale. This class of enzymes usually display a high stereoselectivity and a favourable equilibrium towards amine formation, unfortunately AADH suffers from narrow substrate scope. Enzyme engineering is often employed to improve the dehydrogenase activity towards the desired substrates with a good number of successful attempts reported in literature.

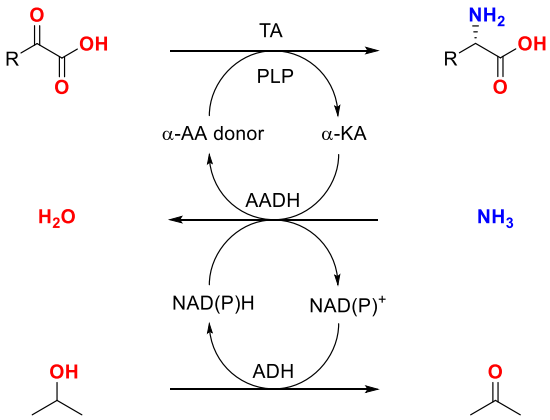
In 2006 Novick[63] evolved the NADPH dependant *C. glutamicum* meso-2,6-D-diaminopimerate dehydrogenase (meso-DAPDH); the resulting quintuple mutant DAADH showed a broad substrate scope and was successfully employed by Turner[7] in the synthesis of D-arylalanines

(Scheme 1.10). Recently Yang *et al.* reported the use of an engineered *Pseudomonas putida* GluDH for the large-scale production of L-PT.[64]



Scheme 1.11 Application of the engineered DAADH for the production of D-Phe analogues. A glucose dehydrogenase (GDH) is coupled in the system for the recycling of NADPH.[7]

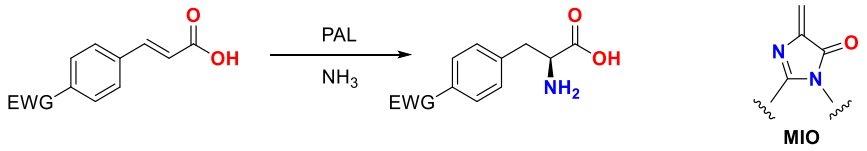
Transaminases (TA) catalyse the PLP mediated transfer of the amine group from a sacrificial amino donor to the substrate. TAs are divided in two main groups: α-TAs, which exclusively bind α-keto acids and ω-TAs which do not require the presence of a carboxylic acid at the α position of the substrate. These enzymes are characterised by a broad substrate scope and high stereoselectivity but an unfavourable thermodynamic equilibrium, which results in poor conversion and difficult product recovery, due to the large excess of amino donors used in the biotransformation. Various strategies employed to overcome this limitation include exploiting product solubility, the use of a smart amino donor easily removed *via* evaporation (isopropylamine),[59] or by filtration of the precipitate polymeric by-product (*o*-xylylenediamine),[65] or the coupling of the TA in an enzymatic cascade for amino donor and cofactor recycling.



Scheme 1.12 Tri-enzymatic cascade for the production of unnatural AAs.[8]

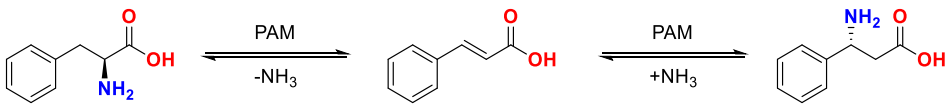
In 2019, the coupling of a TA, an AADH and an alcohol dehydrogenase (ADH) in a tri-enzymatic cascade was successfully employed for the production of unnatural amino acids using a 1:100 of amino donor to substrate ratio (Scheme 1.11).[8]

Ammonia Lyases (AL) and Amminomutases (AM) catalyse the reversible addition of ammonia, or amines, to the C-C double bond of α,β-unsatured acids. The reaction is mediated by the MIO cofactor, which is formed *in situ* in the enzyme active site *via* cyclization and dehydration of highly conserved Ala, Ser and Gly residues; no cofactor recycling strategies are required for these enzymes. The Turner group has carried out extensive research on the phenylalanine ammonia lyase (PAL) enzymes, given their biocatalytic potential and in 2016 reported an engineered variant of *Rhodotorula graminis* PAL for the production of electron-poor L-Phe derivatives (Scheme 1.12).[66]



Scheme 1.13 PAL catalysed reaction and structure of MIO cofactor. [66]

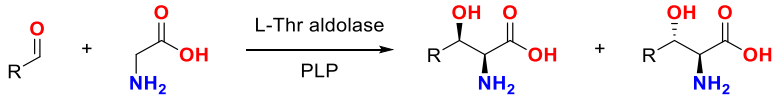
The natural role of AM is to catalyse the isomerization of α to β-AAs, the enzyme is able to transfer the amino group from the α to the β position of the substrate passing through the formation of an α,β-usatured acid as intermediate (Scheme 1.13). If able to control the regioselectivity of the reaction, this class of enzyme shows a high potential for the selective production of β-AAs. Janssen[67] reported an efficient engineered phenylalanine AM for the production of aromatic β-amino acids. Since the reaction pass through the formation of an enolate intermediate, the regioselectivity of the process was tuned by inserting mutations that stabilize the negative charge on the aromatic ring, favouring the attack of the amino group at β position.



Scheme 1.14 Isomerization of α to β-L-amino acid catalysed by PAM.

Aldolases are PLP dependant enzymes, various members of this family (threonine aldolases and α-methylserine aldolases) are able to catalyse the condensation of an amino acid with an aldehyde to form a β-hydroxy-α-amino acid. This condensation has high potential for synthetic applications since aldolases have generally a good aldehyde scope and during the

reaction a new C-C is formed introducing two stereogenic centre in the product (Scheme 1.14).[68] However, these enzymes suffer from poor diastereoselectivity, especially in the formation of the stereocenter at the β carbon; engineering is often required to improve the selectivity of the biocatalytic process.



Scheme 1.15 Biocatalytic aldol condensation between glycine and an aldehyde for the formation of two β-hydroxy-α-amino acid diastereoisomers.

1.3.3.2 Kinetic resolution (KR) of amino acids

The second strategy exploited for the enzymatic synthesis of amino acids is the kinetic resolution of racemic derivatives; which can be carried out with a wide variety of hydrolytic enzymes.[62] The popularity of this process derives from a favourable reaction equilibrium, shifted towards the formation of the product and the high stereoselectivity of hydrolases, furthermore this family of enzymes do not require the use of expensive cofactors. KRs suffer from poor yields; the maximum conversion of a typical process is 50%, leaving half of the starting material (the undesired enantiomer) unreacted (Figure 1.10 A). However, this limitation can be overcome by introducing an *in-situ* racemization step (Figure 1.10 B) to deplenish the unwanted enantiomer and achieve a maximum theoretical conversion and enantiomeric excess of 100% in a dynamic kinetic resolution (DKR).[69,70] To obtain an efficient DKR a highly enantioselective hydrolysis and a fast racemization equilibrium are required, this last step can be either spontaneous, chemo or enzyme catalysed.

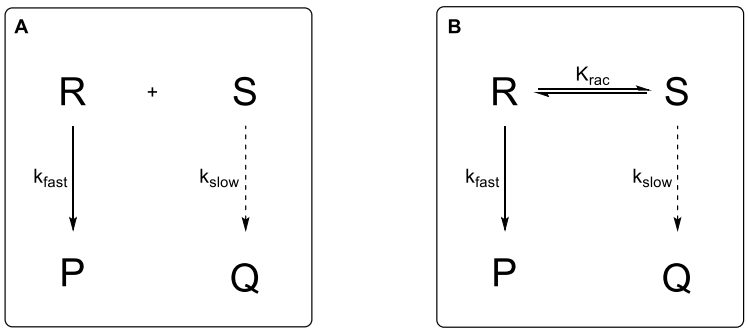
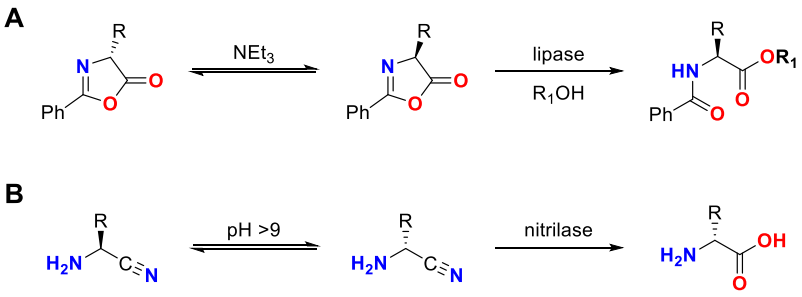


Figure 1.10 Schematic representation of a **A)** KR and **B)** DKR process.

Various chemoenzymatic DKRs have been successfully achieved for the synthesis of alcohols and amines, combining hydrolytic enzymes with Ru and Pd complexes;[71] however, the only example of chemoenzymatic DKR for amino acid preparation was reported by Turner in

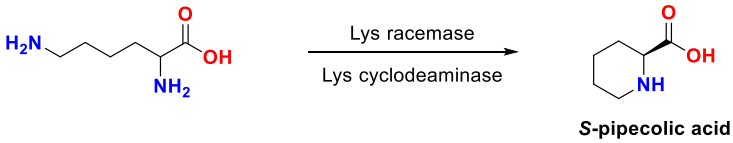
2000.[72] The Novozyme lipase was employed to obtain L-amino acid derivatives (81% yield) from racemic oxazolones, which racemase at alkali pH in the presence of a base (NEt_3), given the low pK_a of the C_4 proton (Scheme 1.15 A). An example of spontaneous racemization was described by Wei *et al.* by employing a nitrilase stable at basic pH to obtain the DKR of phenylglycinonitrile; α-aminonitriles undergoes spontaneous racemization under alkali conditions (Scheme 1.15 B).[73] Chemocatalysed or spontaneous racemisations are not popular strategies, since the majority of enzymes do not tolerate harsh conditions or the presence of transition metal complexes in solution.



Scheme 1.16 Examples of chemoenzymatic DKR. **A)** Chemoenzymatic DKR for the synthesis of AAs derivative from oxazolones. **B)** Spontaneous racemization, followed by enzymatic hydrolysis of α-aminonitriles.[72,73]

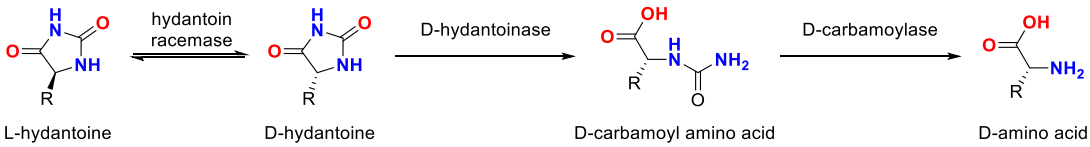
A class of enzymes called racemases (E.C. 5.1), which catalyse the inversion of stereochemistry of a chiral centre of the substrate, have been employed in the development of various biocatalytic DKR.[74] An attractive trait of racemases is their high substrate chemoselectivity, since they bind exclusively to either free amino acids (AA racemase class) or their derivatives; an important factor in a DKR process, where a chemoselective racemization step is required to avoid low product optical purity.

Amino acid racemase (**AAR**) have found application in the synthesis of cyclic amino acid; in this DKR an AAR is coupled with a stereoselective cyclodeaminase (Scheme 1.16).[75] While hydantoin racemase, α-amino-ε-caprolactam (ACL) racemase and *N*-acetyl amino acid racemase (NAAAR) were employed in hydrolytic DKR processes for the production of both natural and non-natural amino acids.[74,76]



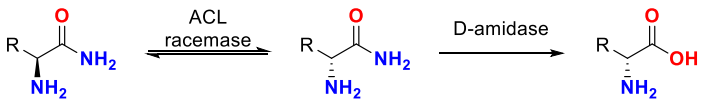
Scheme 1.17 DKR of racemic lysine for the production of *S*-pipecolic acid. [75]

Hydantoin racemase are coupled with two enantioselective hydrolases in the so-called hydantoinase process (Scheme 1.17); in this tri-enzymatic process the two enantiomers of monosubstituted racemic 5-hydantoines, are in a dynamic equilibrium, catalysed by the racemase enzyme. The D-hydantoinase opens the heterocycle *via* hydrolysis of one amide bond and the *N*-carbamoyl-D-amino acid intermediate is then converted into the final product by by the D-selective carbamoylase. This process is employed for the production of D-amino acids given the higher number of D-stereospecific hydantoinases isolated.[77–80]



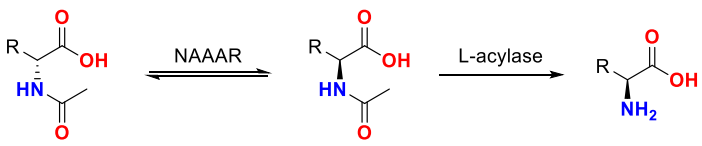
Scheme 1.18 The hydantoinase process; a hydantoin racemase catalyse the equilibrium between L and D-hydantoines, while two enantioselective hydrolytic enzymes (D-hydantoinase and D-carbamoylase) convert the starting material to D-AA.

An **ACL racemase** was also employed in the synthesis of D-arylalanines *via* a biocatalytic DKR of α-aminoamides, by coupling the racemase with a D-selective amidase (Scheme 1.18); unfortunately, ACL racemase substrate scope is usually limited and enzyme engineering is required to improve the biocatalyst efficiency, as reported by Asano.[81]



Scheme 1.19 Biocatalytic DKR of α-amino amides for the production of D-amino acids. [81]

NAAARs are particularly attractive since they possess a broad substrate scope and can be efficiently coupled with acylases in a DKR for the production of α-amino acids (Scheme 1.19).[76] Acylases are highly stereoselective metalloenzymes, which catalyse the hydrolysis of the acetate group of *N*-acetyl amino acids. They are reported to be active for a wide range of *N*-acetyl-NCAAs,[82,83] furthermore both L and D-stereospecific acylases have been isolated by various bacterial strains,[84,85] thus this process can be employed for the synthesis of both D and L-NCAAs.



Scheme 1.20 DKR for the production of L-amino acids from *N*-acetylated derivatives *via* coupling of NAAAR with an L-acylase.

1.3.4 Future challenges

The last decade has been characterized by a growth of the biocatalysis field, with numerous papers and patents being published every year. It is clear that biocatalysis has established itself as a fundamental tool for the production of both high (pharmaceuticals) and low-cost (bulk chemicals) products in small and large scale; however various bottlenecks still plague the sector. One of the main limitations of biocatalysis is the long time and great workload required to evolve enzymes, without any assurance to achieve the desired goal; the development of artificial intelligence and new technology for faster high throughput screening would probably play an important role to speed up the process of evolution.[86]

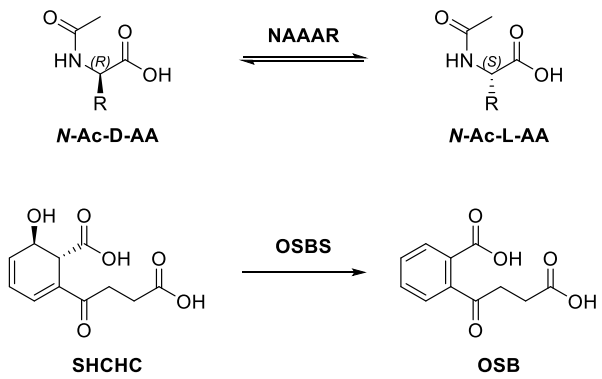
While enzymes which catalyse functional group interconversion are well characterised, more research is required to develop efficient biocatalysts for C-C and C-X functionalization, to extend the repertoire of enzyme catalysed reactions.[57] Moreover, novel enzyme immobilization and flow technologies, along with reactor design will be fundamental to achieve cost-effective batch or continuous biocatalytic process at industrial scale.[86]

1.4 *N*-Acetyl-Amino Acid Racemase

1.4.1 Overview

The discovery of NAAAR dates back at the end of the 20th century, with Tokuyama reporting the isolation of a novel racemase from the actinomycetes *Streptomyces* Y53,[87] that is able to catalyse the stereoinversion of *N*-acetyl-amino acids. Over 49000 microorganisms were screened, however NAAAR activity was exclusively found in actinomycetes strains, such as *Amycolaptopsis* sp. TS1-60,[88] *A. azurea*[89] and *A. orientalis* subsp *lurida*.[90] NAAARs were also isolated from several extremophiles strains: *Deinococcus Radiodurans*,[91] *Thermus thermophilus*,[92] *Geobacillus kaustophilus*[93,94] and *Geobacillus stearothermophilus*.[95]

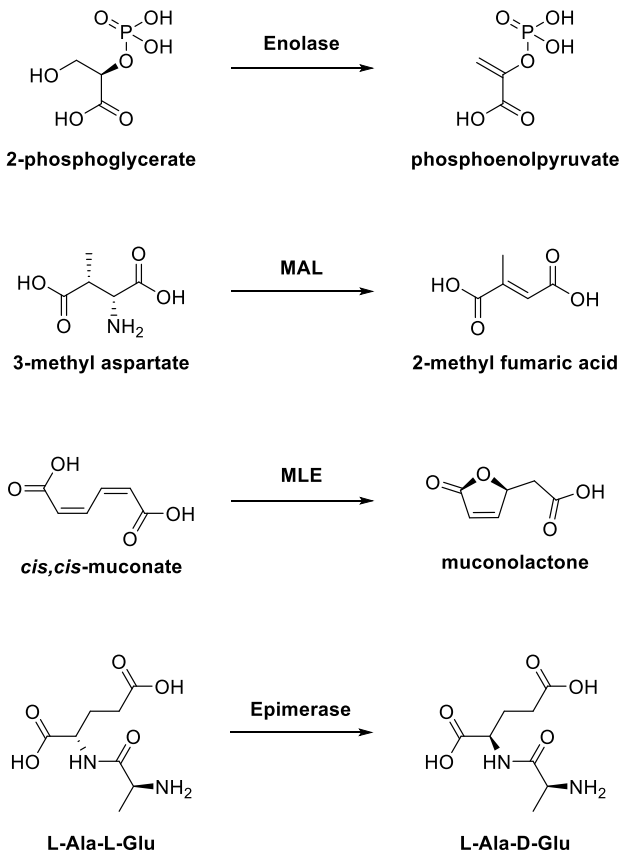
At the time, the production of optically pure amino acids *via* enzymatic KR, was already a well-established process. Various industries expressed great interest in this new racemase class, given their potential in the development of an efficient DKR of *N*-Ac-AAs. Extensive studies were conducted on NAAAR and numerous patents, were published in the last 30 years, as summarised in this work.[96] Unfortunately, the isolated enzymes often exhibit low activity towards *N*-acetylated amino acids and substrate inhibition at concentration > 50 mM, hence, as previously reported, research on biocatalysis focused more on developing efficient methods for the direct asymmetric synthesis of NCAAs.



Scheme 1.21 The bifunctional NAAAR is able to catalyse the racemization of *N*-acyl amino acid (top reaction) and the *syn* dehydration of SHCHC (bottom reaction), an intermediate of menaquinone (vitamin K12).

The lack of activity of NAAAR towards *N*-Ac-AAs was rationalised by Gerlt,[97] he proposed that the racemization of acylated amino acids was not the "correct" function of this class of enzymes. Indeed, NAAAR shows a degree of promiscuity, being able to catalyse two different reactions: the stereoinversion of *N*-acetyl-AAs with a k_{cat} of 10 s^{-1} and the *syn* dehydration of 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) to form *o*-succinylbenzoate

(OSB) with higher efficiency, a k_{cat} of 120 s^{-1} (Scheme 1.20). The SHCHC *syn* dehydration reaction is typically catalysed in bacteria by OSB synthase (OSBS) enzymes; however, no *osbs* gene was found in *Amycopatosis* genome, suggesting that NAAAR was actually the enzyme involved in OSB production in these microorganisms, which would explain the higher activity of the enzyme toward the OSBS reaction. Furthermore, the racemase preference towards *N*-succinylated substrates, with the isomerization of these substrates proceeding 10-fold faster than that of *N*-acetyl-AAs, lead Gerlt[97] to conclude that NAAAR is a bifunctional member of the enolase superfamily, possessing both OSBS and *N*-succinyl-amino acid racemase (NSAAR) activity. To avoid confusion, in this work the term NAAAR will be used to denote the bifunctional enzymes belonging to the OSBS/NSAAR subgroup.



Scheme 1.22 Examples of reactions catalysed by various members of the enolase superfamily. From the top dehydration of 2-phosphoglycerate, 1,2-ammonia elimination of 3-methyl aspartate, lactonization and racemization of L-Ala-L-Glu dipeptide.

1.4.2 The enolase superfamily

The enolase superfamily consists of over 8000 members, these enzymes are characterised by a highly conserved three-dimensional structure despite having diverse amino acid sequences and mechanisms (Scheme 1.21).[77]

The structural domains conserved in the enolase superfamily are an α+β capping domain (Figure 1.11 A and B) involved in substrate recognition and a $(\beta/\alpha)_7\beta$-barrel domain, a variation of the common $(\beta/\alpha)_8$-barrel (TIM-barrel) motif, found in ~10% of the crystal structures published in the protein data bank (PDB). The $(\beta/\alpha)_7\beta$-barrel (Figure 1.11 A and C) contains the catalytic acid/base and metal binding residues which form the active site of these enzymes.[97,98]

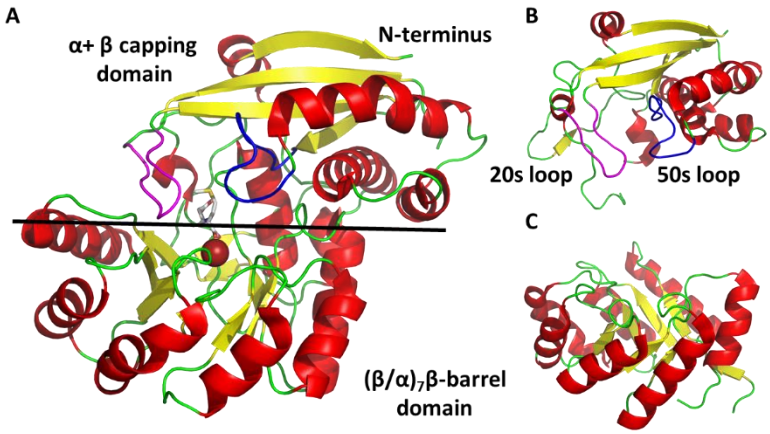
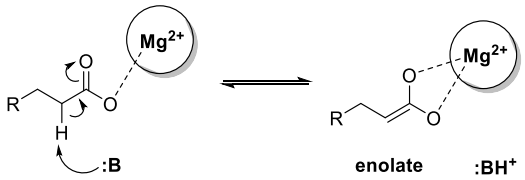


Figure 1.11 The *A.* sp TS1-60 NAAAR crystal structure (1SJC) in complex with Mg^{2+} (red sphere) and *N*-succinyl methionine (grey). **A)** Structure of NAAAR monomer divided in two section; **B)** α+β capping domain (upper section) containing the 20s (magenta) and 50s (blue) loops and **C)** the $(\beta/\alpha)_7\beta$-barrel domain (lower section).

Members of the superfamily are able to catalyse a wide range of reactions, such as racemization, β-elimination, lactone formation, and *syn* dehydration. However, all these biotransformations start with a common step: the formation of an enolate intermediate with the abstraction of the alpha-proton of the substrate (Scheme 1.22).[99] This mechanism also suggests a key role for a metal ion (Mg^{2+}) to stabilize the enolate *via* coordination and have substrates with a C-H in alpha position in a carboxylate group, which decrease the pK_a of the α-proton and stabilise the intermediate *via* resonance. The deprotonation step is carried out by an amino acid residue which acts as general acid/base catalysts, these residues are not highly conserved in the family, which is actually divided in various subgroups based on the number and nature of catalytic amino acids employed for the enolate formation. Mandelate racemases (MR) use a Lys and His, while enolases use only a Lys and muconate lactonizing enzymes (MLE II) use two Lys residues.



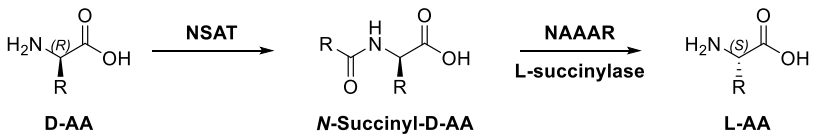
Scheme 1.23 Mechanism of base-mediated and metal-assisted α-proton abstraction of carboxylic acid substrates employed by the members of the enolase superfamily.

Given their conserved architectural motifs and mechanistic diversity, the enolase superfamily was extensively studied by the Gerlt group in order to understand how enzymes can acquire new functions in nature during evolution.[98,100] According to the divergent evolution theory, all members of this superfamily derive from a common ancestor with the ability to abstract the α-proton of carboxylic substrates. This ancestor then acquired new functions *via* selection of mutant variants; with the advantageous mutations being passed down various generations, eventually forming the enolase superfamily we know today. Evidence of this process occurring can be found within the family itself, since its members possess a certain degree of "flexibility" being able to acquire or lose a particular function with just a slight variation in their amino acid sequences. By inserting a single point mutation in the sequence of *E. coli* epimesare AEE and *Pseudomonas* sp. P51 MLE II, two mutants (AEE D297G and MLE II E323G), able to catalyse both their natural reaction and the *syn* dehydration of SHCHC were isolated and characterised. To be noted that the two wild type (WT) enzymes did not show any activity for the OSBS reaction.[101] Furthermore, the OSBS activity of *E. coli* D297G AEE was improved by 5-folds inserting a second mutation (I19F).[102] While D297 and E323 are located on the $(\beta/\alpha)_7\beta$-barrel, I19 is part of the capping domain; however, all three residues are involved in substrate recognition and binding, suggesting that changing substrate specificity is the first step taken during divergent evolution to allow the development of novel enzyme functions. A similar example was recently reported by Glasner *et al.*; where a mutation at the Y299 position of *Alicyclobacillus acidocaldarius* OSBS resulted in the isolation of a bifunctional enzyme with both OSBS and NSAAR activity.[103]

1.4.3 Role, structure and mechanism

The recognised biological role of NAAAR is to catalyse the formation of OSB, the fourth step in the biopathway for the synthesis of menaquinone (vitamin K12, Scheme 1.20).[104] However, it is still not clear why some OSBS enzymes acquired additional NSAAR activity through evolution. Analysis of the genome of the microorganisms containing NSAAR/OSBSs revealed that the gene of this bifunctional enzymes is usually incorporated in an operon containing a

N-succinyl D-amino acid transferase (NSAT) and a M20 hydrolase.[105] These three enzymes may work together to form a pathway for the detoxification of D-amino acids; which are first succinylated by NSAT and then converted to *N*-succinyl-L-AAs by the bifunctional NAAAR; in the last step the hydrolase remove the succinyl group and release the free L-amino acids (Scheme 1.23).[96] Another possible role for NAAAR is to intervene in the synthesis of D-AAs, since these compounds are fundamental in bacteria, as an essential component of peptidoglycans, which contribute to cell wall formations.[106]



Scheme 1.24 Proposed tri-enzymatic pathway for the detoxification of D-amino acids in microorganisms encoding a NSAT and a NSAAR/OSBS bifunctional enzyme.

The three-dimensional structure of NAAAR closely resembles that of the other members of the enolase superfamily; the central $(\beta/\alpha)_7\beta$-barrel domain is highly conversed in the family and contains the two catalytic residues, Lys 163 and Lys263 for *A.* sp. TS1-60 NAAAR, located at the opposite sides of the active site (Figure 1.12 A). The D189, E214 and D239 metal binding residues can also be found in the barrel domain. Following the coordination of the substrate carboxylate to the metal cofactor (Mg^{2+}) the ε NH_2-group of highly conserved lysines act as general acid/base catalysts for the abstraction of the proton from the α-C to generate the planar enolate intermediate. This intermediate can be protonated on the same face by Lys163 or on the opposite face by Lys263, to form the second stereoisomer in a "mirror-like" mechanism (Scheme 1.24).[91,107]

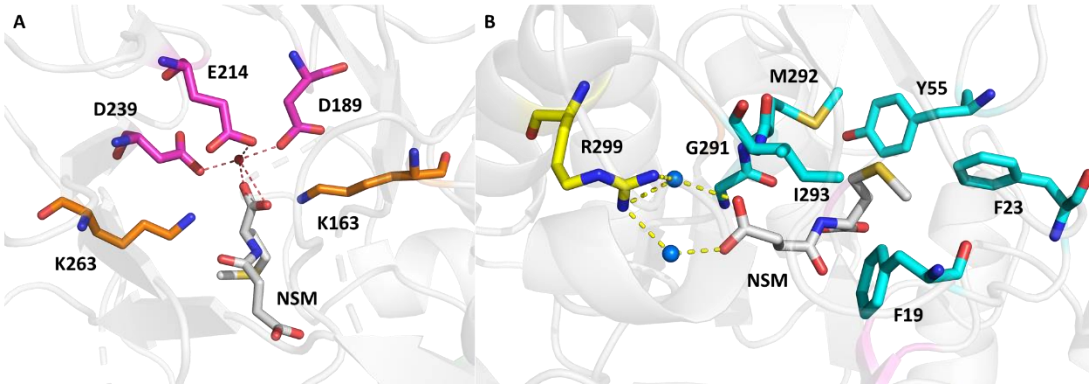
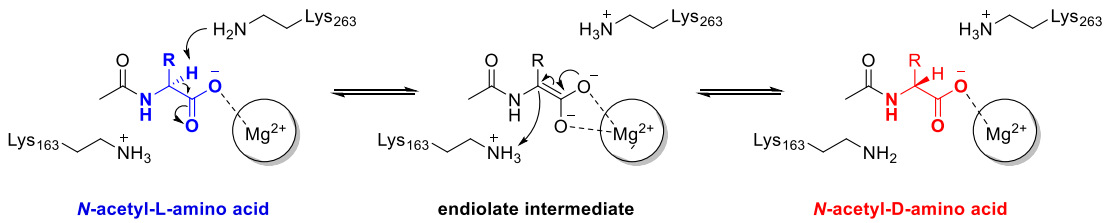


Figure 1.12 *A.* sp TS1-60 NAAAR crystal structure (1JSC) in complex with *N*-succinyl methionine **NSM** (grey) and magnesium (red sphere). **A)** NAAAR active site with the two catalytic lysine K163 and K263 (orange) and Mg(II) binding residues D189, E214 and D239 (magenta). **B)** Substrate binding pocket; R299 (yellow) interacts with the carboxylate of the succinyl group *via* H-bonds formation with two

water molecules, while residues F19, F23, Y55 of the capping domain and G291, M292 and I293 of the $(\beta/\alpha)_7\beta$-barrel create a hydrophobic pocket to accommodate the substrate side chain.

The succinyl side chain of *N*-succinyl-AAs interacts with the R299 residue *via* formation of bridging H-bonds with two molecules of water (Figure 1.12 B). Furthermore, the substrate bound crystal structure (PDB entry code 1SJC) of these enzymes shows two disordered loops in the α+β-capping domain, formed by residues 18-26 (20s) and 45-55 (50s), which seem to play a role in the recognition of *N*-acyl amino acid side chains.[107] When the enzyme is free in solution, the two loops are disordered; however, upon addition of the substrate, these residues close around the molecule shielding it from the solvent. The aromatic or aliphatic AAs side chain upon binding to the enzyme, is surrounded by the residues F19, F23, Y55, G291, M292 and I293, which form a hydrophobic pocket (Figure 1.12 B). While the 50s loop is well conserved in the enolase superfamily, the 20s loop is quite diverse.[100] However, the relevance of its role for the activity of OSBS/NSAAR bifunctional enzymes has been recently proved by Glasner *et. al.*[108] Introducing mutations at position F19, R20, S22 and F23; the catalytic efficiency of *A.* sp. TS1-60 NAAAR was greatly reduced; interestingly the S22R and F23A mutations had a different effect on the two reactions. The arginine mutant showed higher activity for the OSBS reaction, while the NSAAR catalytic efficiency was improved in the F23A mutant. Generally, changes in the 20s loop region affected more markedly the K_M of the racemization reaction, suggesting that the residues of the 20s loop are fundamental for OSBS/NSAAR activity and likely mutations in this region have played a role during the evolution of the OSBS family.



Scheme 1.25 Mechanism employed by NAAAR for the racemization of *N*-acyl amino acids.

1.4.4 Application

The classic application of NAAAR in the synthesis of α-amino acids *via* DKR is a well-established process, which take advantage of the use of enzymes with wide substrate scopes. These racemases are able to bind various *N*-acylated amino acids without showing particular preference for the L or D isomer and being able to accept a range of *N*-acyl moieties, such as acetyl, chloro-acetyl, propionyl, butyryl, formyl, benzoyl and succinyl; however, their activity

is affected by the nature of the AA side chain with low tolerance for the polarity from charged groups.[88] The thermostable *A*. sp. TS1-60 NAAAR was particularly promising as coupling partner of an acylase enzyme, since it retains complete activity up to a temperature of 60°C. Tokuyama co-immobilised this racemase, along with either an L- or D-selective aminoacylase, on DEAE-Toyopearl columns to create two bioreactor for the efficient continuous synthesis of L and D-methionine respectively.[109] Various NAAARs were employed in the development of similar DKRs[80], an example is the use of an overexpressed NAAAR and L-acylase from *D. radiodurans* for the production of L-homophenylalanine (HPA) in a whole cell process, which yielded L-HPA with a 99% *ee*.[110] Unfortunately, the low activity and stability of the L-acylase prevents an efficient re-use of the recombinant cells, with a loss of 65% of the total activity by the third cycle. Several patents describing the application of these racemases for the production of optically pure α-amino acids have also being published.[111–120]

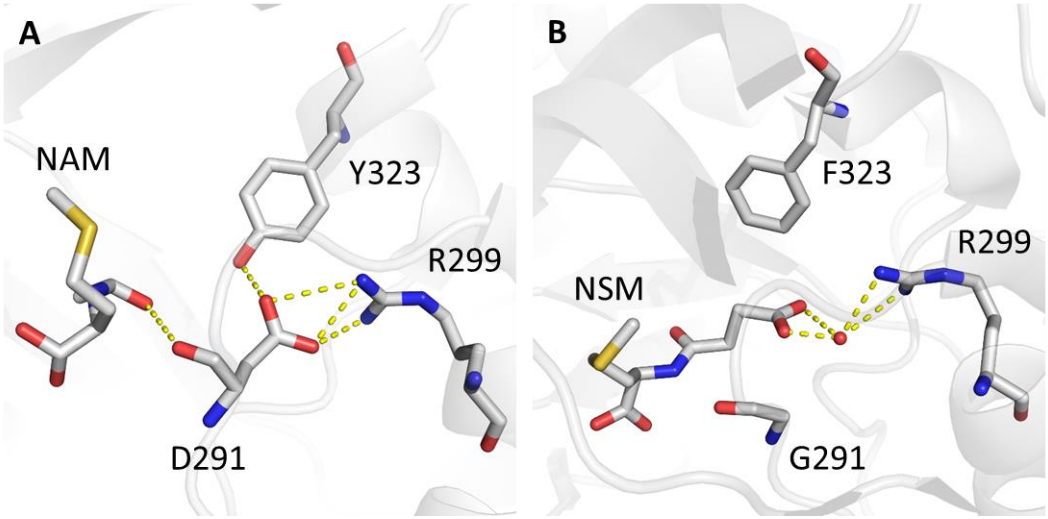
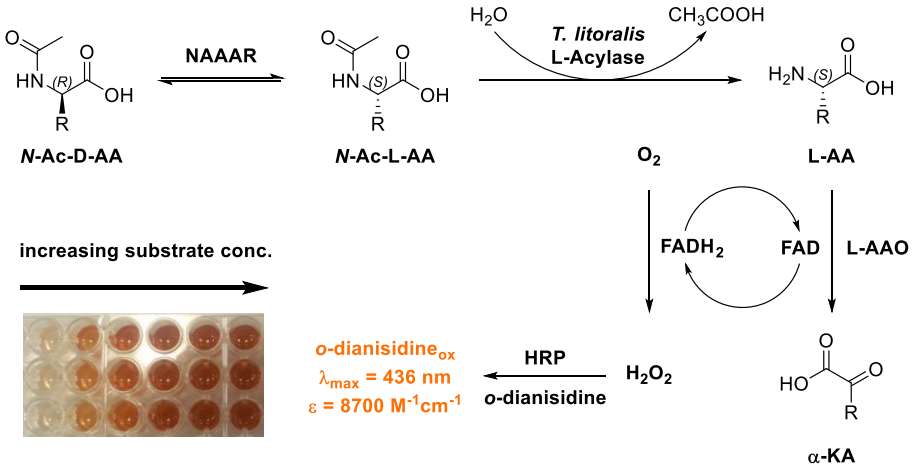


Figure 1.13 Crystal structure of the acyl binding pocket of **A)** NAAAR G291D F323Y (4A6G) in complex with *N*-acetyl methionine (**NAM**). H_2O and H-bonds represented as red spheres and yellow dashed lines respectively and **B)** NAAAR WT (1SCJ) in complex with *N*-succinyl methionine (**NSM**).[121]

The low activity of NAAARs towards *N*-acetylated substrate remains the main limitation of these processes. However, enzyme engineering can be employed to overcome this lack of catalytic efficiency. Baxter *et al.*[121] reported in 2012 the isolation of the *A*. sp. TS1-60 NAAAR G291D F323Y double mutant, characterised by a 6-fold enhancement in activity for various *N*-acetylated canonical and non-canonical amino acids compared to the wild type enzyme. Directed evolution along with an *in vivo* selection system, based on a methionine auxotroph *E. coli* strain, were employed to engineer NAAAR; the improved variant was successfully

employed with a D-acylase in the DKR of D-allylglycine in large scale (50 g/L) with 89% isolated yield and 99% *ee*.

The enhancement of NAAAR activity was rationalised by resolving the NAAAR G291D F323Y crystal structure in complex with *N*-acetyl methionine (PDB entry code 4A6G). The formation of two new hydrogen bond interactions was observed in the acyl binding pocket of the enzyme between the β-carboxylate of D291 and the *p*-OH of Y323 and between the β-carboxylate of D291 and the guanidinium group of R299 (Figure 1.13 A). These new interactions mimic the bridging H-bonds formed by the R299 side chain and the water molecule with the carboxylate group of the succinyl moiety, found in the WT structure in complex with *N*-succinyl-methionine (Figure 1.13 B). However, the acetate moiety lacks this acidic group, resulting in the lower activity of NAAAR WT towards acetylated amino acids.

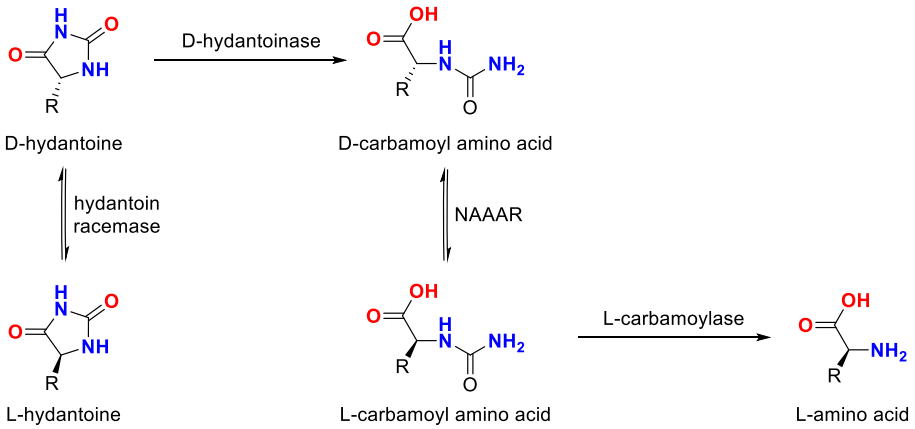


Scheme 1.26 Continuous high-throughput coupled L-AAO assay for the screening of NAAAR activity towards *N*-acetyl canonical and non-canonical amino acids. Picture of an assay plate with constant enzymes loadings and increasing substrate concentration.[122]

The potential for NAAAR evolution was limited in Baxter's work[121] by the use of a selection system based on a bacterial strain with a methionine deficiency; only NAAAR variants with improved activity towards *N*-Ac-D-Met could be isolated using this system. To obtain a more efficient biocatalyst, which could be employed in the synthesis of a wide range of NCAAs, a more generic screening protocol is required. A second work published by the Campopiano group[122] covers the development of a high-throughput continuous colorimetric assay which couples the NAAAR and L-acylase with a FAD dependant L-amino acid oxidase (AAO) to produce hydrogen peroxide in the reaction mixture, which can be easily detected *via* oxidation of *o*-dianisidine, catalysed by the horseradish peroxidase. The oxidised form of *o*-dianisidine possess a dark orange/red colour in solution (Scheme 1.25) with a strong

absorbance at 436 nm (8700 $M^{-1}cm^{-1}$). This robust continuous assay can be employed to measure NAAAR variants activity for various *N*-acetyl NCAAs.

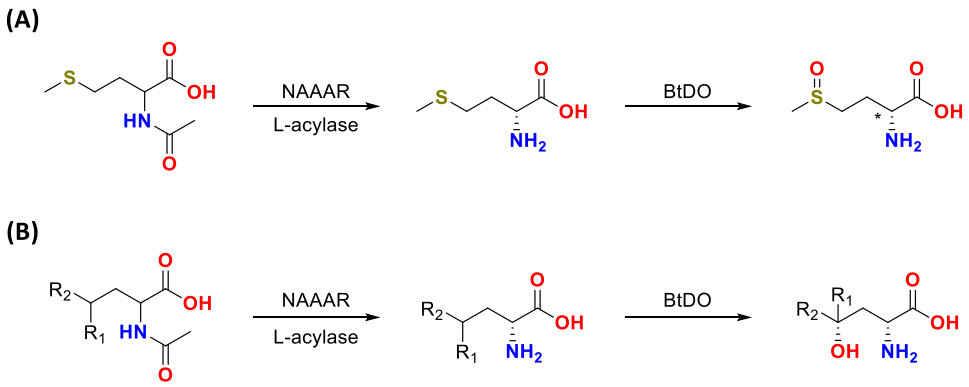
Variants of the classic acylase DKR process were developed by coupling NAAAR with various hydrolase enzymes. The *Geobacillus stearothermophilus* NAAAR was coupled with the *Cupriavidus* sp. P4-10-C *N*-succinyl-D-amino acid desuccinylase for the synthesis of D-phenylalanine and D-tryptophan.[123,124] *N*-carbamoyl-L-amino acids amidohydrolases[125] have also been employed in these kind of processes. *D. radiodurans* NAAAR and *Bacillus kaustophilus* L-carbamoylase were co-expressed in *E. coli* cells; following permealisation with toluene treatments,[126] these cells were employed in a whole-cell DKR process for the production of L-HPA. Recovery and recycling of the biocatalyst was extremely efficient, with a conversion > 90% retained after eight cycles. A similar process for the biocatalytic synthesis of L-HPA was described by Lin *et al.*,[127] where the *D. radiodurans* NAAAR and *B. kaustophilus* L-carbamoylase were immobilised and packed in a bioreactor, for continuous flow production; however, the low stability of the amidohydrolase complicates the process. Unfortunately, large excess of the carbamoylase was required to extend the reactor operational time. More successful was the application of the *G. kaustophilus* NAAAR/ *G. stearothermphilus* L-carbamoylase couple for the production of L-aromatic and aliphatic amino acids starting from racemic *N*-carbamoyl and *N*-formyl derivatives.[128,129] The co-immobilised enzymes were able to yield an average of 80.5% pure L-norleucine, with only a decrease of 3% in activity *per* cycle (75% overall activity maintained after 10 cycles).



Scheme 1.27 The double racemase hydantoinase process for L-amino acid production, obtained by coupling two racemase (hydantoin racemase and NAAAR) and two hydrolytic enzymes with opposite stereospecificity (D-hydantoinase and L-carbamoylase). [78,128,129]

NAAARs applications have expanded in the last five years by employing the racemase in multiple enzyme cascades. As previously mentioned, the "hydantoinase process" is a well-established protocol for the production of D-amino acids, given the prevalent isolation of D-selective hydantoinase;[79,130] however, this classic process was recently re-designed by coupling the *G. kaustophilus* NAAAR with a hydantoin racemase, a D-hydantoinase and an L-carbamoyl amino acid amidohydrolase, in a four-enzymatic double racemase DKR (Scheme 1.26).[78] In this revised system, after formation, the D-carbamoyl amino acid intermediate undergoes racemization, catalysed by the promiscuous NAAAR, followed by hydrolysis by the L-enantioselective carbamoylase to form pure L-AAs. This double racemase hydantoinase process was successfully employed for the preparation of various L-amino acids: norleucine, norvaline, 2-amminobutarric acid and HPA. To scale up the process the four proteins were efficiently immobilised, retaining their stability and activity; unfortunately, intermediate accumulation and incomplete hydrolysis lead to a maximum conversion of 64% when the immobilised enzymes were coupled together.[131,132]

A second interesting tri-enzymatic cascade was reported by Kourist;[133] where the classic DKR of *N*-Ac-AAs, catalysed by engineered NAAAR G291D F323Y and L-acylase, was coupled with a stereoselective leucine dioxygenase from *Bacillus thuringiensis* (Scheme 1.27). This family of dioxygenases are Fe(II)/α-ketoglutarate dependant enzymes which catalyse the oxidation at the γ-position of amino acids. Starting from a cheap racemic *N*-acetylated derivative, this system can produce highly valuable hydroxy and oxo-amino acids with high diastereomeric excess (*de*). Operating under optimal conditions *S,S*-γ-hydroxy leucine, norleucine and norvaline were synthesised and L-methionine-(*S*)-sulfoxide was isolated in 97% yield and 95% *de*.



Scheme 1.28 Tri-enzymatic cascade for the synthesis of diastereopure **A)** (S,S)-methionine sulfoxide and **B)** (S,S)-γ-hydroxy amino acids.[133]

2 Aims

The overall aim of this work is the development of an efficient DKR system for the biocatalytic production of L-phosphinothricin and other NCAAs. The first and second sections focus on the application of two hydrolytic enzymes, for the synthesis of L-amino acids from *N*-acetylated derivatives.

1) *E. coli N*-acetyl-L-ornithine deacetylase (ArgE).

- Expression and purification of *E. coli* ArgE.
- Enzyme characterization.
- Kinetic analysis.
- Quantitative assay development for reaction monitoring.
- Substrate screening.
- Structural and mechanistic studies.
- Enzyme engineering.

2) *S. viridochromogenes N*-Ac-PTT deacetylase (*Sv*Dea).

- Expression and purification of *S. viridochromogenes* Dea.
- Enzyme characterization.
- Kinetic analysis.
- Substrate screening.
- Structural studies.

The third section will focus on the coupling of the acylase enzymes with engineered NAAAR variants and the optimization and scale up of the biocatalytic DKR process.

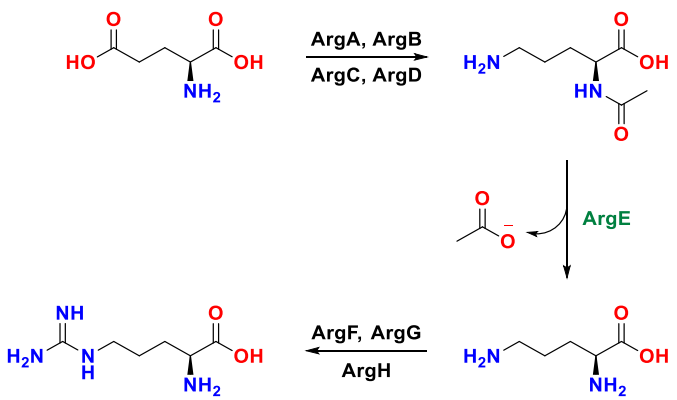
3) *Amycolaptopsis* sp TS1-60 *N*-acetyl amino acid racemase (NAAAR).

- Expression and purification of *A.* sp TS1-60 NAAAR.
- Kinetic analysis.
- Coupling of NAAAR and aminoacylase in a DKR.
- Substrate screening.
- Scale up reactions.

3 The *N*-acetyl-L-ornithine deacetylase ArgE

Currently, the *Thermococcus litoralis* L-Acylase (E.C. 3.5.1.14)[134–136], developed by Littlechild and colleagues, is used in combination with *Amycolaptopsis* NAAAR[121] for the dynamic kinetic resolutions (DKR) of amino acids. However, the intellectual property (IP) surrounding this enzyme is held by Dr Reddy's Laboratories and not available for a PhD project sponsored by Syngenta (Jealott's Hill, UK). Furthermore, while the L-Acylase has shown activity towards aliphatic and aromatic *N*-acetylated amino acids, poor to no activity was reported towards charged and polar substrates.[134] In order to develop an efficient DKR for the production of L-phosphinothricin (PT), a new L-Acylase (amide hydrolase), that can be efficiently coupled with *Amy* NAAAR, needs to be identified and characterised.

The *Escherichia coli N*-acetyl L-ornithine deacetylase (ArgE) was reported to be active towards *N*-Ac-L-PT in a patent from Bartsch *et al.*[137] Indeed, it was successfully used for the creation of various transgenic plants with male sterility including tobacco[138] and rice.[139] By expressing the *argE* gene in the tapetum tissues of these organisms pollen maturation was disrupted. In another application the gene was introduced in perennial ryegrass[29] to create an efficient and selective elimination system for this weed, to reduce its growth in golf courses and tennis courts. These results reported in the literature strongly suggest that ArgE is an ideal candidate for the DKR of L-PT, having a complementary substrate scope compared to Dr. Reddy's L-Acylase, which displays greater activity for hydrophobic *N*-acetylated derivatives. [134]



Scheme 3.1 Bacterial biosynthetic pathway to L-arginine. ArgE (in green) catalyses the fifth step, the deacetylation of *N*-acetyl-ornithine.

The enzyme named ArgE, first isolated from *E. coli* in 1992,[140] is a zinc dependant metallo-protein, that is member of the M20A peptidase superfamily (EC 3.5.1.16).[141] This class of enzymes catalyses the fifth step (acetylornithine hydrolysis) in the bacterial biosynthesis of arginine from glutamic acid (Scheme 3.1).[142] The intermediate L-ornithine is of fundamental importance for these organisms, since it used not only for the production of L-Arg but also in the synthesis of polyamines; making the ArgE class a target for the production of new inhibitors.[143]

3.1 Aims

The aims of this project were to express, purify and characterize *E. coli* ArgE and to assay its activity and enantioselectivity for the resolution of L-PT and other non-canonical amino acids (NCAAs). We also planned to optimize the coupling of ArgE with NAAAR to develop a DKR. To achieve this goal a quantitative assay for monitoring the deacetylation reaction was optimised. Furthermore, to elucidate the mechanism of the enzyme and investigate which residues are involved in substrate binding, the crystal structure of ArgE requires to be resolved.

3.2 Expression and purification of un-tagged ArgE.

The full length *argE* gene, cloned into the expression vector pET28a using the NcoI and XhoI restriction sites, was ordered from GenScript (Uniprot code: P23908). The un-tagged pET28a-ArgE construct was transformed in *E. coli* BL21 (DE3) cells and a series of expression tests were carried out to optimise the incubation temperature and isopropyl β-D-1-thiogalactopyranoside (IPTG) concentration used for the enzyme production. All samples were analysed *via* SDS-PAGE electrophoresis. The gels show a low yield of soluble protein, with the majority of ArgE expressed in the inclusion bodies at 45 kDa (Figure 3.1 A and B).

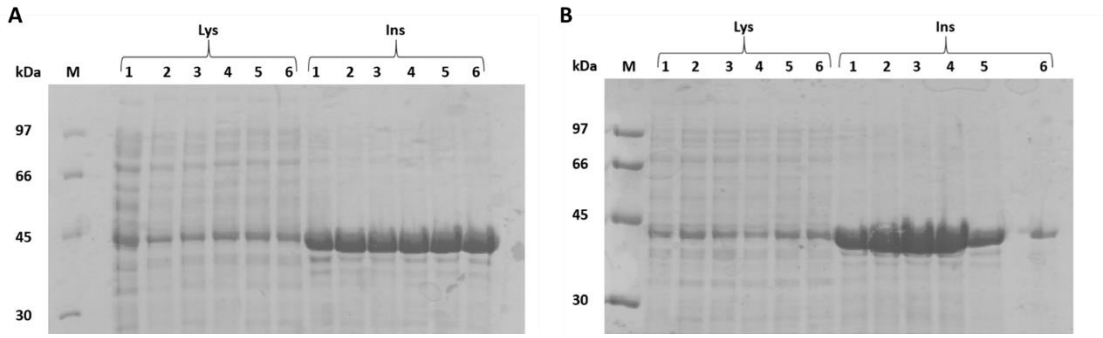


Figure 3.1 12% SDS-PAGE gel of un-tagged ArgE expression tests. Lane 1 low molecular weight marker LMWM (M), lane 2-7 soluble (Lys) and 8-13 insoluble (Ins) fractions of samples 1 (0.0 mM IPTG), 2 (0.1 mM IPTG), 3 (0.2 mM IPTG), 4 (0.4 mM IPTG), 5 (0.8 mM IPTG) and 6 (1.0 mM IPTG), incubated **A)** at 20°C for 20 h and **B)** at 30 °C for 4 h.

Since ArgE is a zinc dependant metallo-enzyme, it was thought that the addition of Zn(II) in the media may increase the yield of soluble protein. A second series of expression tests were carried out by adding various concentrations of $ZnSO_4$ in the samples. The SDS-PAGE gels show a general increase in the yield of soluble ArgE (Figure 3.2) and the best conditions selected for scale-up were: 0.1 mM $ZnSO_4$, 0.2 mM IPTG at 20°C for 20 h.

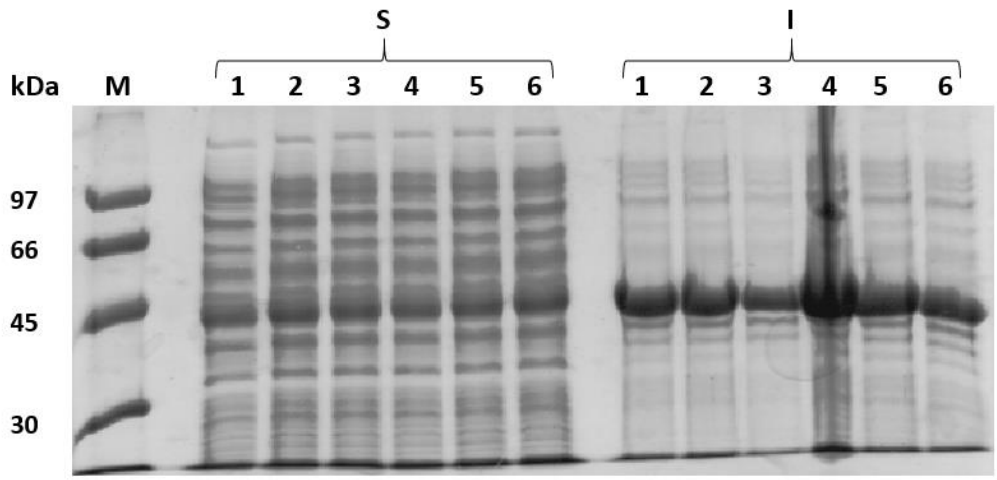


Figure 3.2 12% SDS-PAGE gel of un-tagged ArgE expression tests. Lane 1 LMWM (M), lane 2-13 soluble (S) and insoluble (I) fractions of samples 1 (0.0 mM $ZnSO_4$), 2 (50 µM $ZnSO_4$), 3 (0.1 mM $ZnSO_4$), 4 (0.25 mM $ZnSO_4$), 5 (0.5 mM $ZnSO_4$) and 6 (1.0 mM $ZnSO_4$), incubated at 20°C for 20 h with 0.2 mM IPTG.

Un-tagged ArgE was purified following a two-step protocol; the cell free extracts (CFEs) were applied to an anion exchange Q Sepharose Fast Flow (QFF) column with the enzyme eluting at ~300 mM NaCl (0.1-1.0 M NaCl gradient). These fractions were collected and concentrated before separated on a pre-calibrated Sephacryl S300 size exclusion chromatography (SEC)

column. The gel filtration chromatogram shows two main peaks at 40 mL and 65 mL respectively, however SDS-PAGE analysis confirms that the enzyme (band at ~42 kDa) eluted on the shoulder of the second peak at 73 mL (Figure 3.3 A and B), which corresponds to a molecular weight (MW) of 85 kDa for native ArgE (homodimer). After purification, the un-tagged ArgE purity was ~70% (estimated by gel electrophoresis), unsuitable for mass spectrometry analysis and further applications. To optimize the purification, protocol the *arge* gene was subcloned to obtain a histidine tagged construct.

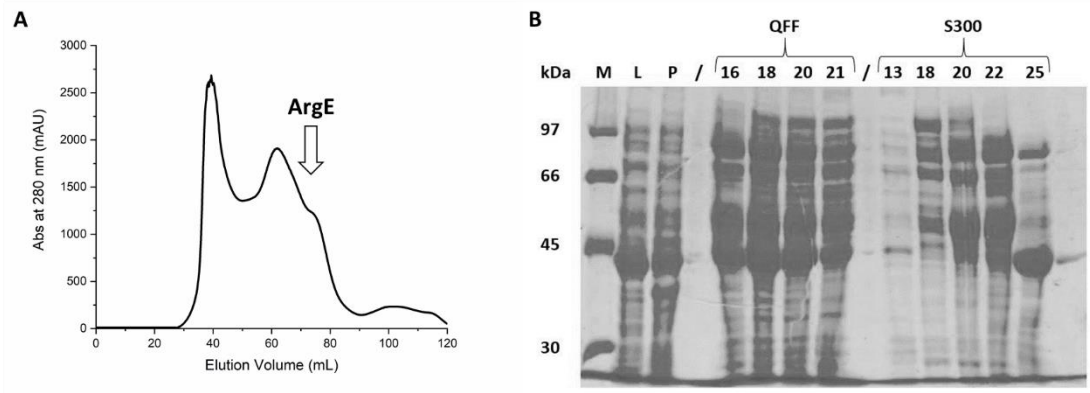


Figure 3.3 Untagged *E. coli* ArgE purification. **A)** Size exclusion chromatogram for un-tagged ArgE on a Sephacryl S300 column. The arrow points to the acylase peak. **B)** 12% SDS-PAGE gel of un-tagged ArgE purification: lane 1 LMWM (M), lane 2 CFE (L), lane 3 insoluble fraction (P), lane 4-7 QFF fractions (5 mL each) and lane 8-12 S300 fractions (3 mL each). Untagged ArgE eluted in fraction 25 at 73 mL elution volume, on the shoulder of the second peak.

3.3 Expression and purification of N-HisTag ArgE

The *arge* gene was sub-cloned into a pETHISTEV plasmid, using the NcoI/XhoI restriction sites, to obtain a construct with an N-terminal tobacco etch virus (TEV) cleavable hexahistidine tag. The protein was expressed using the optimised protocol and purified *via* metal ion affinity chromatography (IMAC) with a HisTrap Fast Flow 1 mL column, followed by SEC (Superdex S200 column). The SEC absorbance chromatogram, registered at 280 nm, shows a single peak at an elution volume of 75 mL (Figure 3.4 A), which corresponds to a molecular weight of 90 kDa for the homodimeric 6xHisTag ArgE.

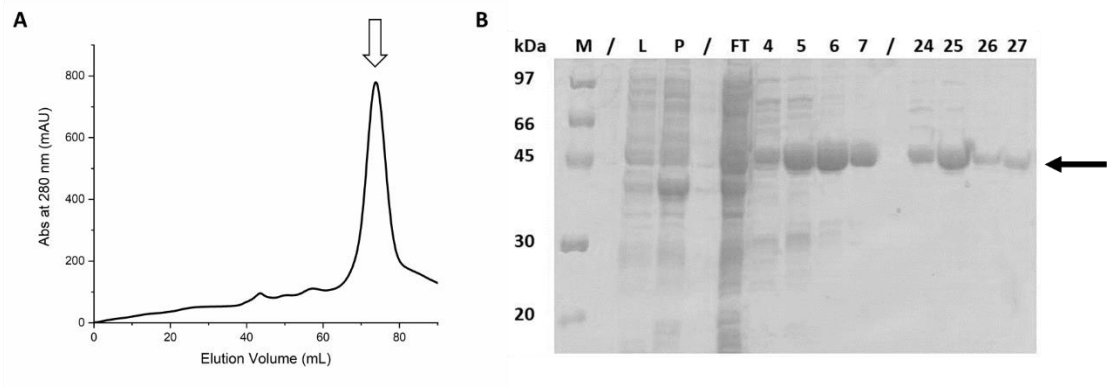


Figure 3.4 Purification of N-His tagged ArgE. **A)** Size exclusion chromatogram for tagged ArgE on a Superdex S200 column. The arrow points to the acylase peak. **B)** 12% SDS-PAGE gel of N-HisTag ArgE purification: lane 1 LMWM (M), lane 2 CFE (L), lane 3 insoluble fraction (P), lane 4 HisTrap flow through (FT), lane 5-8 HisTrap fractions (3 mL each, enzyme elution start at ~200 mM imidazole) and lane 9-12 S200 fractions (3 mL each).

After purification ArgE was obtained with a purity of ~95% (Figure 3.4 B) and an average yield of 15 mg *per* L of culture. To confirm that the correct construct was expressed and purified, the molecular weight of ArgE was calculated using denaturing liquid chromatography electrospray-mass spectrometry (LC ESI-MS). The resulting spectrum displayed several peaks, representing the many charge states of the protein. The observed MW of 45369.00 ± 0.90 Da (Figure 3.5 B), was in accordance with the theorical MW minus the initial methionine: 45369.50 Da, calculated from the protein sequence with the ExPaSy ProtParam tool (https://web.expasy.org/protparam). The loss of the initial methionine is a very common post-translational modification.[144]

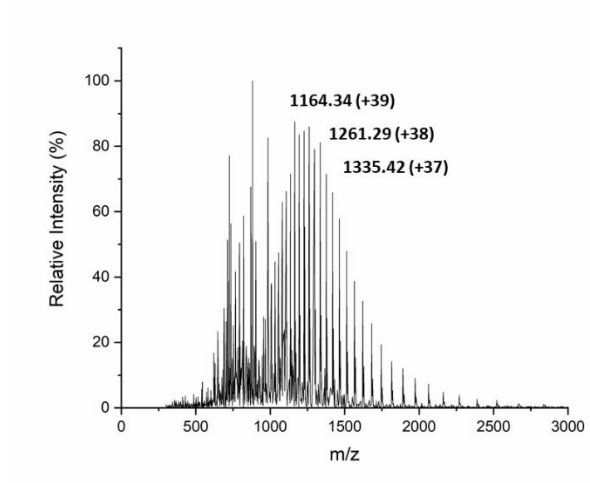
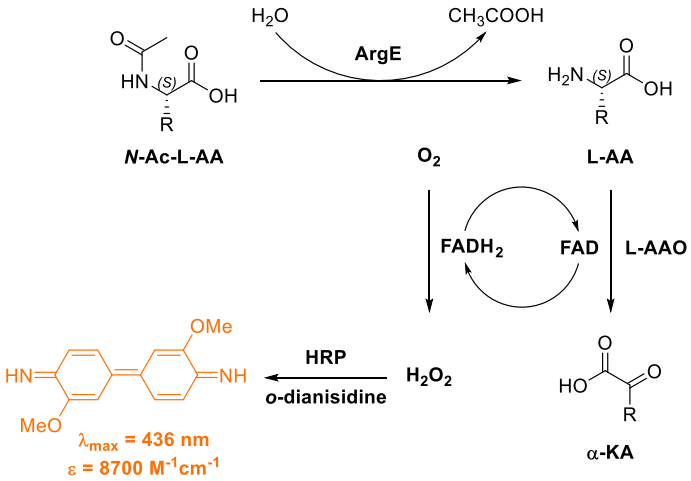


Figure 3.5 Denaturating LC ESI-MS of ArgE (20 µM) with a mass of 45369.00 ± 0.90 Da, in accordance with the MW calculated from the sequence (45369.50 Da). The values reported are *m/z* with the charges in brackets.

3.3.1 High-throughput colorimetric L-amino acid oxidase assay

The high-throughput continuous L-amino acid oxidase (L-AAO) assay reported by Campopiano *et. al.*[122] was employed for the kinetic analysis of ArgE (Scheme 3.2). The *N*-Ac-L-AA substrate is converted by the acylase to the L-AA; which is oxidised by the commercial *Crotalus adamanteus* L-AAO (lyophilised venom from Sigma-Aldrich) to form the corresponding α-keto acid (α-KA), plus ammonia. The L-AAO is a flavoprotein and during the oxidation FAD is reduced to $FADH_2$, which is promptly re-oxidised by oxygen, producing hydrogen peroxide (H_2O_2) as a reaction by-product. Finally, H_2O_2 reacts with horseradish peroxidase (HRP) in the presence of *o*-dianisidine, which in its oxidised form has a strong red/orange colour and absorbance at 436 nm (ε = 8700 $M^{-1}cm^{-1}$), easily monitored spectrophotometrically.



Scheme 3.2 Reaction scheme of the ArgE/L-AAO continuous assay. The oxidised from of *o*-dianisidine has a strong absorbance at 436 nm and can be easily monitored *via* UV-Vis on a 96-wells microplate reader.

The assay was employed to measure the kinetic parameters of ArgE towards *N*-Ac-L-Met, which is reported to be the best substrate for this acylase.[141] A set of reactions at increasing substrate concentrations were incubated at 40°C and the absorbance of *o*-dianisidine$_{(ox)}$ was monitored at 436 nm every 30 s for 1 h. The initial reaction rates were calculated from the linear portion of the absorbance curves and fitted using the Michaelis-Menten non-linear fit with Lab Origins software (Figure 3.6 A and B). The calculated kinetic parameters show that ArgE has a strong affinity (K_M = 2.4 ± 0.2 mM) and high turnover number (TON, k_{cat} = 4.58 ± 0.07 s^{-1}) for *N*-Ac-L-Met. Hence, this enzyme has potential for biocatalytic applications, as suggested by its high catalytic efficiency (k_{cat}/ K_M = 1759.653 ± 0.001 $M^{-1}s^{-1}$).

Similar results have already been reported by Blanchard *et. al.*,[141] however the catalytic constant reported in literature (1700 ± 40 s^{-1}) deviates highly from the value calculated with the L-AAO assay. This discrepancy can be explained by the difference in assays employed to assess the enzyme activity. Blanchard *et al.*[141] performed the kinetic analysis by directly monitoring the change in the absorbance of the amide bond (ε = 103 $M^{-1}cm^{-1}$). Alternately, the L-AAO assay relies on a cascade of reactions to couple the deacetylation step with the formation of *o*-dianisidine$_{(ox)}$ (ε = 8700 $M^{-1}cm^{-1}$).[122]

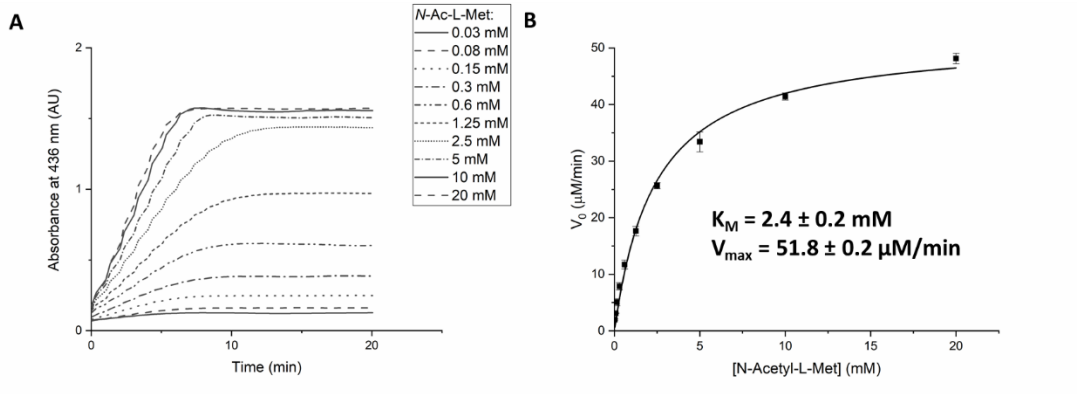


Figure 3.6 Determination of *E. coli* ArgE kinetic parameters *via* L-AAO continuos assay. **A)** Typical raw data profile for the formation of *o*-dianisidine$_{ox}$ over time at various substrate (*N*-Ac-L-Met) concentrations. **B)** Michaelis-Menten plot of ArgE for *N*-Ac-L-Met. Initial rates calculated from the linear portion of the absorbance curves and fitted with the OriginLab software. The reaction mixture contained: *N*-Ac-L-Met (0-20 mM), ArgE 10 µg/mL, $CoCl_2$ 100 µM, L-AAO 70 mU/mL, HRP 10 U/mL and *o*-dianisidine 0.1 mg/mL in NaPi buffer 0.1 M pH 8. The plate was incubated at 40°C for 1 h and absorbance measured at 436 nm every 30 s. Reactions repeated in triplicates, error bars calculated as the standard deviations.

3.3.2 ArgE kinetic analysis

The L-AAO assay proved to be a quick and efficient method to measure enzymatic activity; as reported by Campopiano *et. al.,*[122] the L-AAO used possesses quite a broad substrate scope, being able to accept both natural (methionine or phenylalanine) and non-natural (phenylglycine or 2-naphtylalanine) amino acids. However, the full potential of this oxidase is yet to be explored.

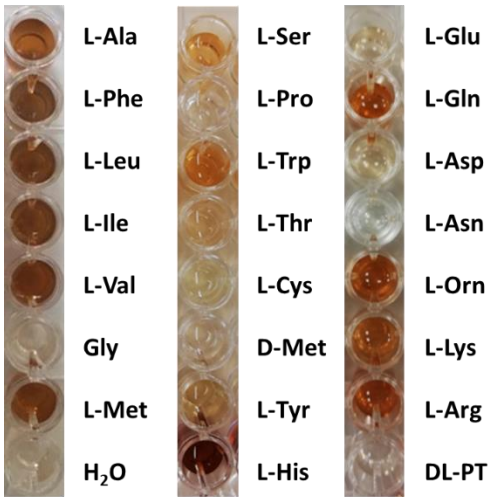


Figure 3.7 Screening of L-AAO activity towards the 20 canonical L-amino acids, plus phosphinothricin (DL-PT) and ornithine (L-Orn). Water and D-methionine used as controls. The assay mixture contained: AA 10 mM, $CoCl_2$ 100 µM, L-AAO 100 mU/mL, HRP 10 U/mL and *o*-dianisidine 0.1 mg/mL in NaPi buffer 0.1 M pH 8. The plate was incubated at 40°C for 1 h and absorbance measured at 436 nm every 30 s.

A series of reactions were set up using the 20 proteogenic L-amino acids, plus phosphinothricin (PT, **2a**) and L-ornithine (L-Orn, **2d**), to determine the substrate range and the optimal concentration of the enzyme to use in the reaction mixture. For this initial activity screening a concentration of AAs of 10 mM and of L-AAO of 100 mU/mL was used.

***N*-Ac:**	**K_M (mM)**	**k_{cat} (s^{-1})**	**k_{cat}/K_M ($M^{-1}s^{-1}$)**
L-Met	2.4 ± 0.2	4.58 ± 0.07	1759.653 ± 0.001
L-Ala	9.91 ± 0.04	0.077 ± 0.002	7.773 ± 0.001
L-Leu	8.3 ± 0.6	1.18 ± 0.02	142.72 ± 0.01
L-Ile	22.1 ± 0.8	0.69 ± 0.03	31.153 ± 0.001
L-Val	60 ± 6	0.107 ± 0.003	1.792 ± 0.002
L-Phe	-	-	-
L-Tyr	N.D.	N.D.	N.D.
L-Trp	N.D.	N.D.	N.D.
L-His	0.96 ± 0.02	0.126 ± 0.003	131.7512 ± 0.0006
L-Arg	24 ± 4	0.070 ± 0.005	2.93 ± 0.001
L-Lys	5.2 ± 0.7	0.185 ± 0.009	36.423 ± 0.006
L-Gln	12.6 ± 0.9	0.65 ± 0.04	51.735 ± 0.004
DL-Ser	15 ± 5	0.011 ± 0.001	0.78 ± 0.04
L-Thr	-	-	-

Table 3.1 Kinetic parameters of ArgE for fourteen *N*-Ac-AAs obtained with the continuous L-AAO assay The reaction mixture contained: *N*-Ac-AA various concentrations, $CoCl_2$ 100 µM, L-AAO 100-300 mU/mL, HRP 10 U/mL and *o*-dianisidine 0.1 mg/mL NaPi buffer 0.1 M pH 8. The plate was incubated at 40°C for 1 h and absorbance measured at 436 nm every 30 s.

As shown in Figure 3.7 the oxidase is active for most hydrophobic and positively charged amino acids, with a few exceptions (proline, glycine and asparagine). The oxidase is also able to accept some polar amino acids (serine and threonine), as substrates, while no activity was detected for: aspartic acid, glutamic acid and PT, it is clear that the L-AAO does not tolerate AAs with negatively charged side chains. Unfortunately, this assay cannot be used to screen for acylase activity towards *N*-Ac-PT (**1a**), however by optimizing the loading of the oxidase, the assay was employed to measure the kinetic parameters of ArgE towards fourteen of the twenty canonical *N*-Ac-L-AAs (Table 3.1).

The substrate scope of the enzyme was wider than anticipated; being able to accept *N*-acetyl-L: lysine, arginine, histidine, valine, threonine and phenylalanine, while no activity was detected towards *N*-acetyl-L-tyrosine and tryptophan. However, the kinetic parameters for *N*-acetyl-L-phenylalanine and threonine could not be determined, since the kinetic curves did not reach saturation, even at high (> 200 mM) substrate concentration. Interestingly, no activity was previously reported towards all these substrates by Blanchard *et. al.*,[141] another advantage of the higher sensitivity of the L-AAO assay.

3.4 Kinetic resolution of *N*-acetyl methionine

To obtain an efficient kinetic resolution or DKR of *N*-acetyl AAs, it is necessary to employ a highly stereoselective acylase. A chiral HPLC assay, previously optimised by Baxter,[121] was employed to test ArgE stereoselectivity towards *N*-Ac-Met. Standards of: *N*-Ac-L-Met, *N*-Ac-D-Met, L-Met and D-Met were run on a Chirobiotic T (Teicoplanin) column with an isocratic gradient of a 0.025% triethylammonium acetate (TEAA) : MeOH = 75:25 mixture as eluent. The four isomers were separated with good retention times: 3.0, 4.3, 5.6 and 6.7 min, respectively (Figure 3.8 A) and four calibration curves were produced to allow quantitative analysis. (Appendix 19).

A kinetic resolution (KR) was set up with 60 mM *N*-Ac-DL-Met, 100 µM $CoCl_2$ and 10 µg/mL of ArgE in Na phosphate buffer (NaPi) 0.1 M pH 8. Given ArgE high turnover number (TON) for this particular substrate, the acylase concentration had to be kept low, to be able to monitor the reaction conversion over time. The biotransformation was incubated at 40°C for 24 h; samples were taken over time and quenched by 1:40 dilution into the eluent mixture and run on the HPLC.

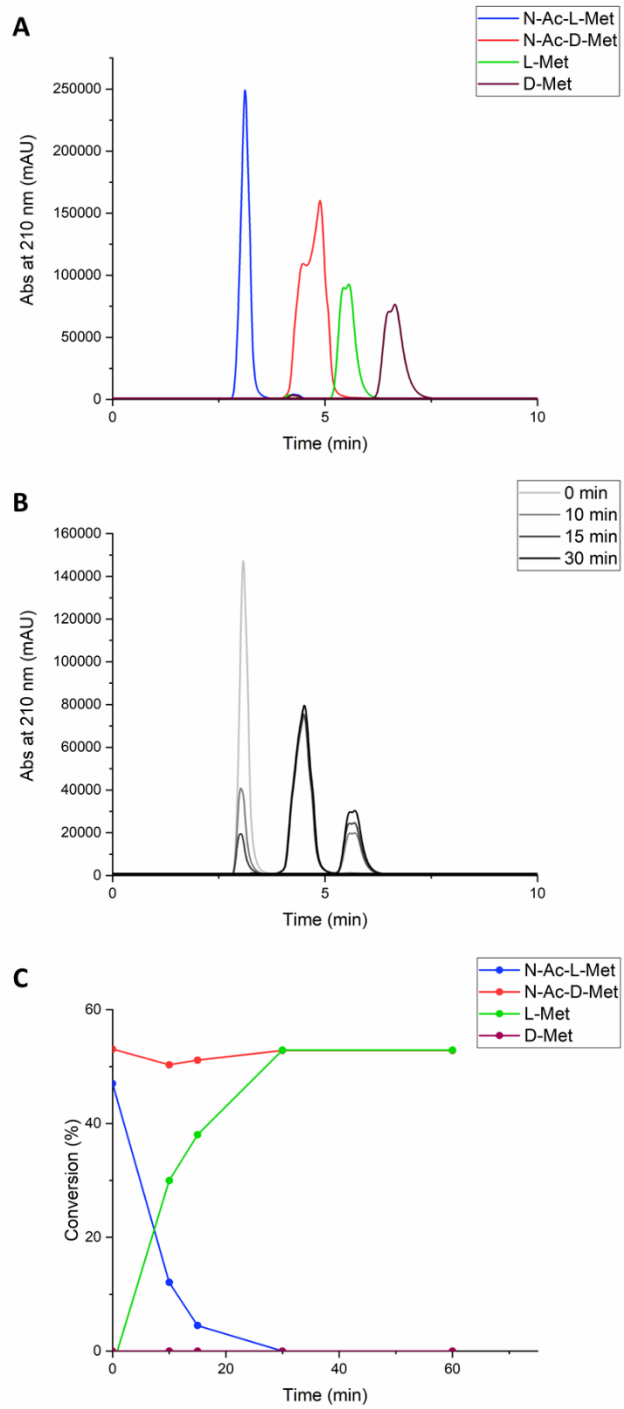


Figure 3.8 KR of *N*-Ac-DL-Met catalysed by ArgE. **A)** HPLC profile for the resolution of both enantiomers of the starting material: *N*-Ac-L-Met (blue), *N*-Ac-D-Met (red) and product: L-Met (green), D-Met (purple). Standards samples at a concentration of 25 mM. **B)** HPLC curves of time point samples of the biocatalytic KR of *N*-acetyl methionine (scale of grey). **C)** Conversion of *N*-Ac-L-Met (blue) into L-Met (green) over time, *N*-Ac-D-Met (red) concentration remains constant over time and no trace formation of D-Met (purple) was detected. All samples analysed by HPLC on a Chirobiotic T (Teicoplanin) column, isocratic gradient: elution phase 0.025% TEAA and MeOH (75/25 ratio), flow 1.0 mL/min, T = 40°C and λ = 210 nm. Reactions repeated in duplicates, error bars calculated as standard deviations reported only for triplicates.

As expected, the reaction was very fast with full conversion (50%) achieved in only 30 minutes with the complete disappearance of the *N*-Ac-L-Met peak at 3.0 min (Figure 3.8 B). Furthermore, ArgE showed complete stereoselectivity towards L-Met formation (*ee* > 99.9%), while *N*-Ac-D-Met concentration remained constant over time and no traces of D-Met formation were observed in the HPLC traces (Figure 3.8 B and C).

3.5 Metal binding studies

3.5.1 Metal content of purified ArgE

ArgE belongs to the di-zinc dependant M20 metallo peptidase family; all members of this family have highly conserved metal-binding residues in their active sites and can accept, as cofactors, zinc or other divalent small transition metal ions, such as Co(II) or Mn(II).[142] Indeed, preliminary studies conducted by Holz *et. al.*[145] suggest that ArgE is able to coordinate up to two Zn^{2+} or Co^{2+} ions *per* monomer (Table 3.2) and both metals cations assume a distorted penta-coordinated geometry, upon binding to the active site residues.

	Zn^{2+}	**Co^{2+}**
k_{d1} (µM)	2.7	0.4
k_{d2} (µM)	51	153

Table 3.2 Dissociation constants (k_{d1} and k_{d2}) for ArgE with Zn^{2+} and Co^{2+} ions, calculated by isothermal titration calorimetry (ITC) by Holz *et. al.*[145]

To determine the metal content of ArgE a sample of purified enzyme (1 mg/mL, ~20 µM) was analysed *via* inductively coupled plasma-mass spectrometry (ICP-MS), while a second ArgE sample (10 mg/mL, ~200 µM) was analysed using a newly developed method reported by Dr. Wong[146] for the determination of the metal/protein stoichiometry *via* Thin-Film Energy Dispersive X-ray Fluorescence Spectroscopy (EDXRF). The EDXRF analysis was run by Silvia Fruncillo, Dr Wong group, University of Manchester. The two analyses gave concordant results: one equivalent of zinc is bound *per* ArgE monomer, while no traces of cobalt were detected; only the [Zn(ArgE)] species is present in solution after protein purification. The data obtained and the dissociation constants (k_{d1} and k_{d2}) values reported in literature (Table 3.2) suggest that while one zinc ion is tightly bound to the enzyme (k_{d1} = 2.7 µM) and probably plays an essential role during catalysis, the second metal binding site (k_{d1} = 51 µM) is probably just partially occupied in physiological condition and has mainly a structural role.

It has been reported that cobalt bound ArgE, [Co(ArgE)] is 2.4 times more active then [Zn(ArgE)];[145] this species is usually prepared by incubation of APO ArgE with an excess of $CoCl_2$. The APO enzyme is obtained by dialysing the purified protein against a buffer containing EDTA. However, this process is quite lengthy and has a poor yield, since a good percentage of enzyme precipitates after $CoCl_2$ addition. To obtain [Co(ArgE)] the enzyme was expressed using the optimised protocol (Chapter 6.3.4) but supplementing the media with 0.1 M of $CoCl_2$, instead of $ZnSO_4$. The purified protein, analysed by IPC-MS, contained 0.5 eqv of zinc bound *per* ArgE monomer: [$Zn(ArgE)_2$], while no traces of Co(II) were detected by the instrument. The enzyme favourably binds to the zinc ions, which are present at low concentration in the LB media, as a consequence half of the protein is expressed in its APO form, which would explain the poor yield (3 mg *per* L of culture) obtained from this batch.

The direct production of [Co(ArgE)] is quite challenging, it may be achieved but the protocol needs optimisation, for instance: by changing the expression media or using another salt as cobalt source. Since the purified [Zn(ArgE)] already presents a high catalytic efficiency, this species was used for future studies.

3.5.2 Effect of Zinc and Cobalt on ArgE activity

To understand if an increase of enzymatic activity can be achieved by addition of metal to the reaction mixture; a kinetic study of: [Zn(ArgE)] and [$Zn(ArgE)_2$] was conducted, employing the continuous L-AAO assay. The kinetic parameters of the two enzyme species were determined by varying the metal content in the assay reaction mixture: $CoCl_2$ 100 µM, $ZnCl_2$ 100 µM and a control without metal addition.

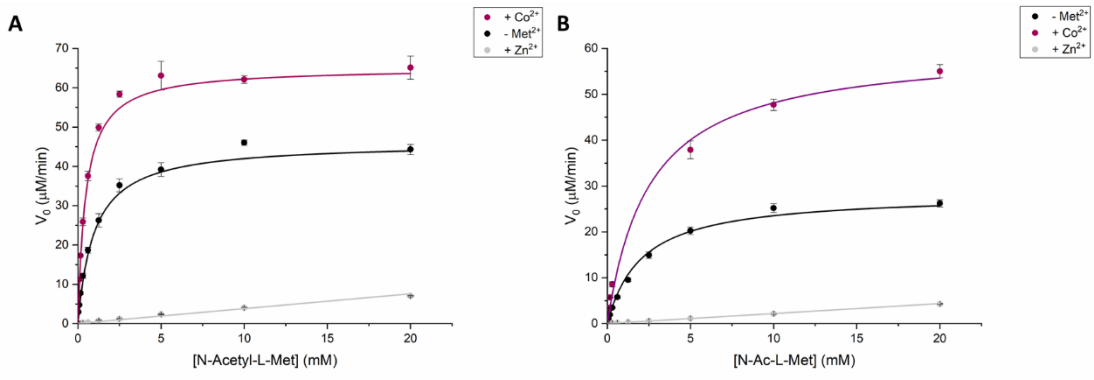


Figure 3.9 Michaelis-Menten plot for the kinetic analysis of **A)** [Zn(ArgE)] and **B)** [$Zn(ArgE)_2$] with *N*-Ac-L-Met in the presence of various metal ions: Co^{2+} (purple), Zn^{2+} (grey) and no metal addition (black). All reactions contained various concentrations of *N*-Ac-L-Met, $CoCl_2$/$ZnCl_2$ 100 µM, L-AAO 70 mU/mL, HRP 10 U/mL and o-dianisidine 0.1 mg/mL in NaPi buffer 0.1 M pH 8. The plate was

incubated at 40°C for 1 h and absorbance measured at 436 nm every 30 s. Reactions repeated in triplicates, error bars calculated as the standard deviations.

As shown in Figure 3.9 A the addition of Co(II) to the reaction mixture caused a slight increase in [Zn(ArgE)] activity (1.4-fold), while an excess of Zn(II) inhibited the enzyme, a phenomenon already observed for other zinc metallo-proteins with a binuclear site.[142] As expected, without any metal addition to the reaction mixture, the activity of [Zn(ArgE)$_2$] is halved compare to [Zn(ArgE)]. However full activity can be regained by addition of Co^{2+} (Figure 3.9 B).

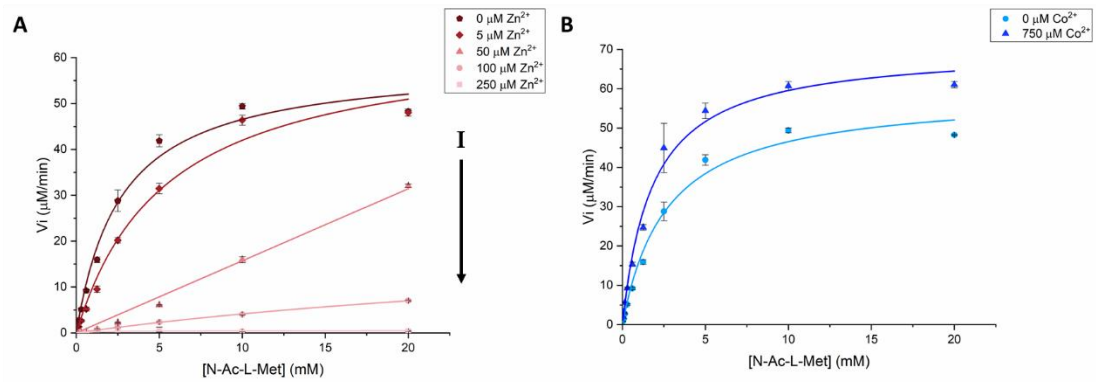


Figure 3.10 Inhibition study of ArgE using **A)** $ZnCl_2$ (scale of red) and **B)** $CoCl_2$ (scale of blue). Each kinetic curve was obtained using a reaction mixture with various concentrations of N-Ac-L-Met, $CoCl_2$or $ZnCl_2$ 0-750 µM, L-AAO 70 mU/mL, HRP 10 U/mL and *o*-dianisidine 0.1 mg/mL in NaPi buffer 0.1 M pH 8. The plate was incubated at 40°C for 1 h and absorbance measured at 436 nm every 30 s. Reactions repeated in triplicates, error bars calculated as the standard deviations.

The inhibitory effect of zinc was further studied by determining the kinetic profile of the ArgE catalysed hydrolysis of *N*-Ac-L-Met at various $ZnCl_2$ concentrations (0, 5, 50, 100, 250 µM). The derived Michaelis-Menten curves (Figure 3.10 A) show a typical competitive reversible inhibition profile, with a constant V_{max} and an increase of the apparent K_M at increasing Zn(II) concentrations. A similar behaviour has been previously reported for the carboxypeptidase A enzyme (M24 family),[147] another zinc dependant hydrolase, active towards ester and amide bonds. Hypothetically, the excess zinc ions could bind to some residues in the active site, which are normally involved in substrate recognition, leaving the Zn^{2+} and substrate to compete for the active site, inhibiting the reaction. This scenario can explain why zinc acts as a competitive inhibitor for the two hydrolases. However, no inhibition was observed even at high $CoCl_2$ concentrations (750 µM, 5 times higher than k_{d2}) in the reaction mixture (Figure 3.10 B).

Since Co^{2+} ions act as ArgE activator, a slight excess of $CoCl_2$ will be used in all future experiments to achieve the highest possible reaction conversion and enzymatic activity.

3.6 ArgE Thermostability

3.6.1 Thermal Denaturation Assay

Thermostable proteins are usually favoured for biocatalytic applications, being easier to handle and use in industrial scale process.[148] To determine the ArgE transition unfolding temperature (melting temperature, T_m) a thermal denaturation assay (TDA) was run on a reverse transcription polymerase chain reaction (RT-PCR) machine. In a typical TDA[149,150] the enzyme is incubated with a fluorophore, as SYPRO-orange[149] (λ_{ex} 470 nm /λ_{em} 570 nm) and the assay temperature is increased constantly over time. The dye fluorescence is quenched by water, but as the protein start to unfold and expose on the surface its hydrophobic residues; SYPRO-orange binds to these regions restoring its fluorescence (Figure 3.11 A).

The assay was run on a 96 well plate and ArgE melting curves (Figure 3.11 A) were recorded in twelve different conditions: 0.1 M Na citrate (pH 3.0, 3.5, 4.0, 4.5, 5.0 and 5.5), 0.1 M NaPi (pH 6.0, 6.5, 7.0, 7.5 and 8.0) and 0.1 Na carbonate (NaCi) pH 9.0. The enzyme's T_ms were calculated from the second derivative of the corresponding melting curves.

As shown in Figure 3.11 B, ArgE thermal stability (T_m ~35°C) is very low at acidic pH, however it significantly increase towards neutral and basic pH values, stabilising at around a T_m of 65°C. Thermostable proteins usually have melting temperatures higher than 80°C, thus ArgE cannot be classified as such. However, it is stable at neutral and basic pH up to a temperature of 60°C, which may increase the chance of protein crystal formation.[151]

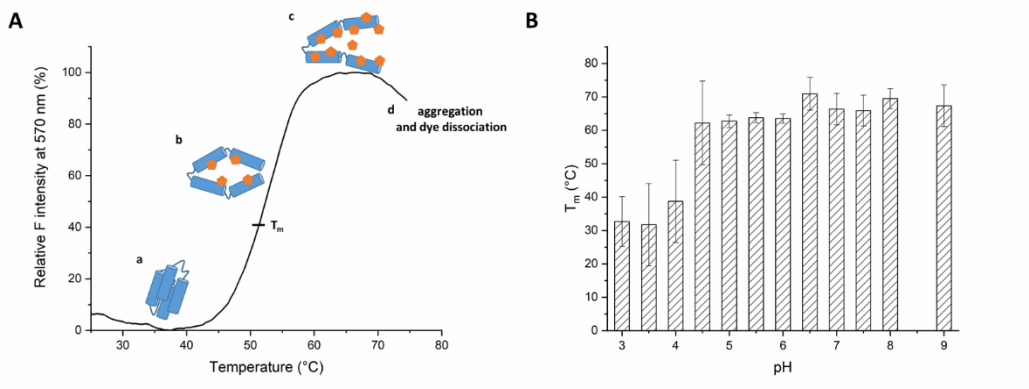


Figure 3.11 Analysis of ArgE thermostability at various pHs *via* TDA. **A)** ArgE melting curve at pH 6.5 (25-75°C). At low temperature the enzyme is stable in solution and SYPRO-orange fluorescence is

completely quenched (**a**), as the temperature increase the protein start to unfold and the dye binds to the hydrophobic residues (**b**). The fluorescence intensity reaches its maximum when the enzyme is completely unfolded (**c**); as the protein starts to aggregate the dye dissociates and the florescence intensity is quenched (**d**). The flex point of the melting curve corresponds to the enzyme T_m. **B)** ArgE melting temperatures at various pH (3-9); the assay mixture contained ArgE 2.0 µM (0.1 mg/mL), 5X SYPRO orange in DMSO and 50 mM buffer at various pHs. The melting curves were obtained measuring the fluorescence of SYPRO orange between 25 and 75°C with a gradient of 1°C/min. Reactions repeated in triplicates, error bars calculated as the standard deviations.

3.6.2 Correlation between reaction temperature and ArgE activity

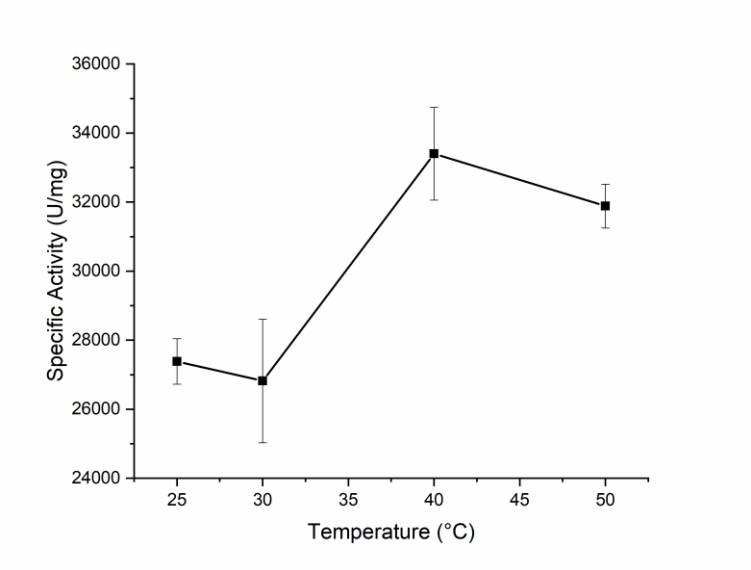
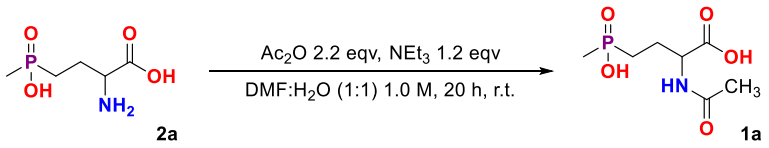


Figure 3.12 ArgE specific activity for *N*-acetyl L-methionine at 25°, 30°, 40° and 50°C measured *via* the continuous L-AAO. The assay mixture contained various concentration of *N*-Ac-L-Met, $CoCl_2$ 100 µM, L-AAO 70 mU/mL, HRP 10 U/mL and *o*-dianisidine 0.1 mg/mL in NaPi 0.1 M pH 8. The plate was incubated at the chosen temperature for 1 h and absorbance measured at 436 nm every 30 s. The specific activity is expressed in Unites *per* milligram of enzyme, with one unit referring to the µM of substrate the enzyme turns over in one minute. Reactions repeated in triplicates, error bars calculated as the standard deviations.

ArgE activity was monitored at four different temperatures; the continuous L-AAO was employed to determine the specific activity of the enzyme for *N*-Ac-L-Met at: 25°, 30°, 40° and 50°C. According to Figure 3.12, ArgE shows high activity (> 25000 U/mg) in this temperature range with an optimal reaction temperature of 40°C (32700 U/mg). Still, the enzyme retained 96% of its activity at 50°C, suggesting that with some optimization it may be possible to couple ArgE with the thermostable NAAAR enzyme in dynamic kinetic resolution.

3.7 ArgE activity towards *N*-Ac-PT

3.7.1 Synthesis and Crystal Structure



Scheme 3.3 Acetylation protocol for DL-phosphinothricin.

The free racemic PT amino acid (gifted by Syngenta, Jealott's Hill, UK), was acetylated with an excess of acetic anhydride in the presence of triethylamine at room temperature (r.t.) over twenty hours (Scheme 3.3). This protocol gave pure *N*-Ac-DL-PT (**1a**) triethylammonium with a yield of 73%.[83]

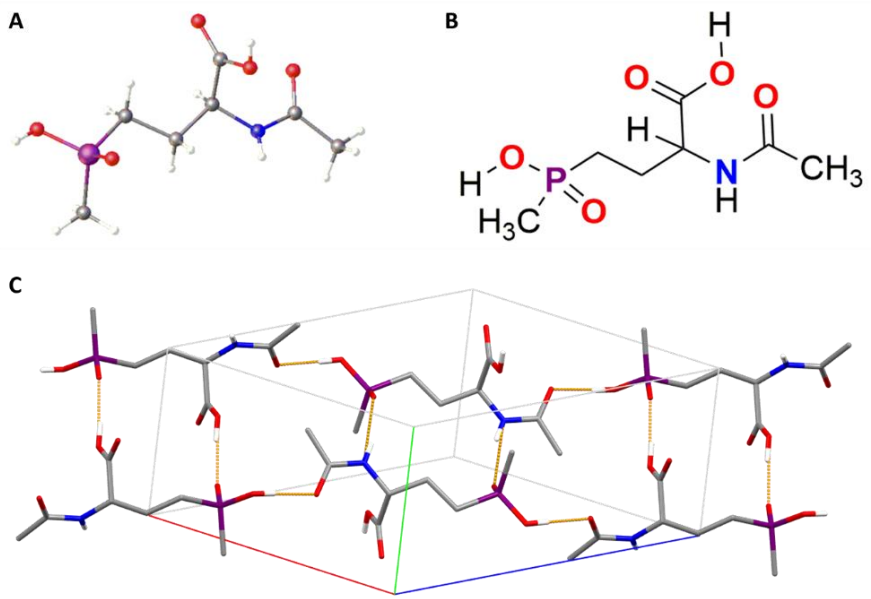


Figure 3.13 *N*-Ac-PT crystallization study. **A)** Resolved X-ray structure and **B)** schematic representation of *N*-Ac-PT (**1a**). **C)** Three-dimensional arrangement inside the unit cell with the formation of intermolecular H-bonding. Carbon in grey, hydrogens in white, oxygens in red, nitrogen in blue, phosphorus in purple and H-bonds in yellow. *N*-Ac-PT molecules (represented as sticks).

A sample of the product was recrystallized by the slow evaporation of a mixture of water and acetone, to give single colourless plate-shaped crystals. The crystals were sent to the X-ray facility of the University of Edinburgh and the X-ray structure was solved with ShelXT (Sheldrick, 2015) structure solution program using the Intrinsic Phasing solution method

(Figure 3.13 A). An extended three-dimensional intermolecular H-bond network was observed (Figure 3.13 B).

3.7.2 Kinetic Resolution of *N*-Ac-PT

The KR of *N*-Ac-PT was set up by incubating the substrate with ArgE and $CoCl_2$ in NaPi 0.1 M pH 8 at 40°C (250 rpm) for 24 h. Eight biotransformations were run simultaneously varying the concentration of the three reaction components, as shown in Table 3.3. The commercially available Amano Acylase and Dr Reddy's *T. litoralis* L-Acylase were also tested. All samples were analysed by LC-TOF MS, but no product formation was detected.

These results were unexpected, since the ArgE catalysed *N*-acetyl PT deacetylation had been previously observed in literature *via* radioactive assay[138] and exploited for creation of transgenic plants.[29,138,139] However, Flashel *et. al.* conducted in 2005[152] a study aimed to understand the stability of the *N*-acetyl PT amide bond under various conditions. The data collected suggests that higher temperature, basic pH and the presence of particular micro-organism in the sample determine an increase in the free PT concentration over time. These results call into question the data described in the literature, regarding the specific activity reported for various L-Acylases, including ArgE, towards this specific substrate.

Enzyme	[*N*-Ac-PT] (mM)	$CoCl_2$ (µM)	L-Acylase (mg/mL)	PT m/z
ArgE	20	100	1.0	N.D.
	20	200	1.0	N.D.
	20	200	5.0	N.D.
	20	200	10	N.D.
	40	200	5.0	N.D.
	40	200	5.0	N.D.
	10	200	5.0	N.D.
***T.litoralis* L-Acylase**	20	200	10	N.D.
Amano Acylase	20	200	10	N.D.

Table 3.3 KR of *N*-Ac-PT, optimization of the substrate, enzyme and cofactor loadings. No product formation detected at the TOF MS.

3.7.3 Inhibition tests

3.7.3.1 Inhibition studies with *N*-Ac-PT and racemic PT

The natural role of L-PT is as a potent inhibitor of the enzyme glutamine synthase in bacteria and plants.[30] It is possible for this amino acid to act in a similar way towards ArgE, with the L-

PT released in solution, binding to the enzyme and inhibiting the reaction, which would explain the low specific activity of ArgE reported for this substrate in the literature.[152]

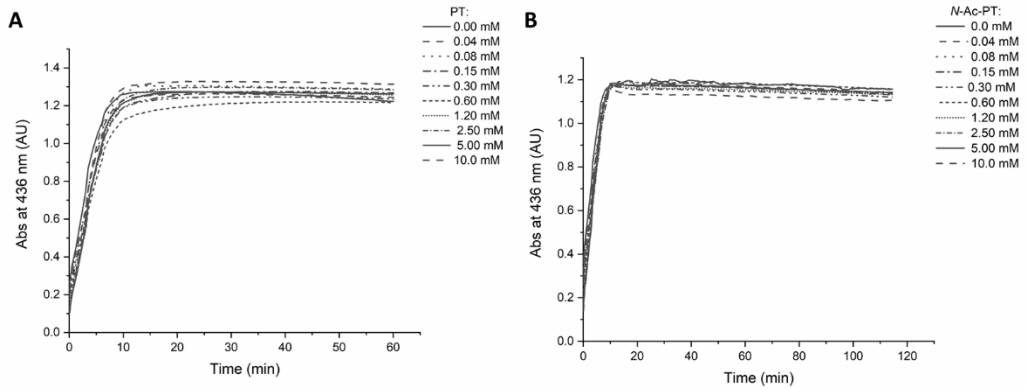


Figure 3.14 Absorbance curves of *o*-dianisidine$_{ox}$ at 436 nm over time. The L-AAO reaction mixture contained *N*-Ac-L-Met 1 mM, ArgE 10 µg/mL, various concentrations of **A)** PT or **B)** *N*-Ac-PT, L-AAO 70 mU/mL, $CoCl_2$ 100 µM, HRP 10 U/mL and *o*-dianisidine 0.1 mg/mL in NaPi buffer 0.1 M pH 8. The reactions were incubated at 40°C and Abs red at 436 nm every 30 s for 2 h.

To test this theory the rate of a set of ArgE-catalysed *N*-Ac-L-Met hydrolysis reactions were monitored, exploiting the high-throughput L-AAO assay (Scheme 3.3); each reaction was incubated with increasing concentrations (0-10 mM) of either PT (**2a)**, or N-Ac-PT (**1a)**. However, no sign of inhibition was detected up to 10 mM concentration of either of these molecules, suggesting that ArgE does not favourably bind these molecules (Figure 3.14 A and B).

3.7.3.2 Inhibition studies with L-methionine sulfoxide and sulfone

To understand why ArgE has such a low binding affinity for *N*-Ac-PT, the same inhibition studies were repeated using, as inhibitor candidates, two L-amino acids structurally analogous to PT: L-methionine sulfoxide (L-MetO, **2b**) and L-methionine sulfone (L-$MetO_2$, **2c**). These three AAs have polar side chains of similar length, furthermore, the sulfoxide, sulfone and phosphonic acids group share a similar sizes and same tetrahedral geometry (Figure3.15).

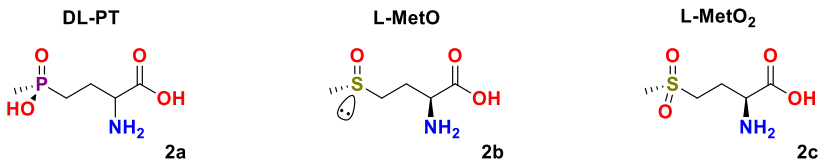


Figure 3.15 Comparison between phosphinothricin (**2a**), L-methionine sulfoxide (L-MetO, **2b**) and L-methionine sulfone (L-$MetO_2$, **2c**) structures.

As shown in Figure 3.16 A and B, neither L-MetO, or L-$MetO_2$ produced an evident inhibitory effect on *N*-Ac-L-Met hydrolysis. However, the reaction control of both assays (which contained 1 mM of **2b** and **2c**, but no substrate) shows a significant background absorbance, suggesting that the L-AAO is active towards the two methionine derivatives. This hypothesis was successfully tested (data not shown) and the L-AAO assay was subsequently used to screen ArgE activity towards these two new potential substrates.

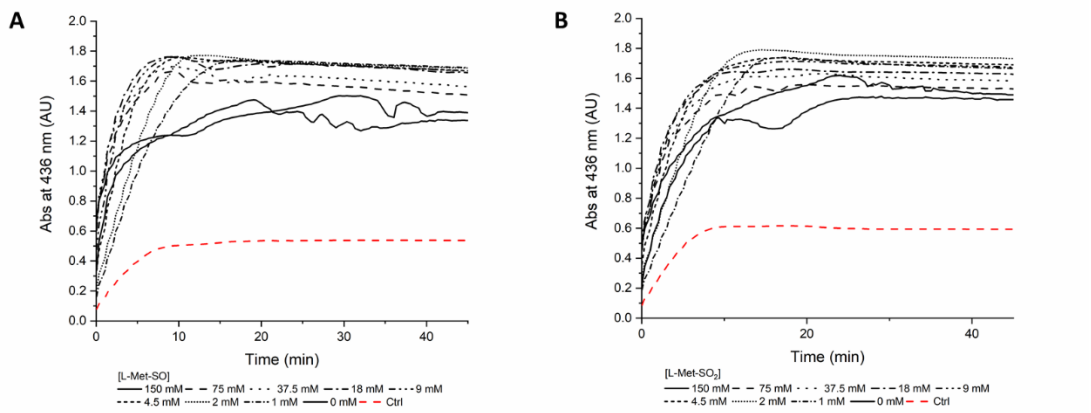


Figure 3.16 Absorbance curves of *o*-dianisidine$_{ox}$ at 436 nm over time. The L-AAO reaction mixture contained *N*-Ac-L-Met 1 mM, ArgE 10 µg/mL, various concentrations of **A)** L-MetO or **B)** L-$MetO_2$, L-AAO 70 mU/mL, $CoCl_2$ 100 µM, HRP 10 U/mL and *o*-dianisidine 0.1 mg/mL in NaPi buffer 0.1 M pH 8. The reactions were incubated at 40°C and Abs red at 436 nm every 30 s for 2 h. Reaction curves in black, control in red.

N-Ac-L-MetO (**1b**) and *N*-Ac-L-$MetO_2$ (**1c**) show a high affinity for ArgE, with a k_{cat} of 1.97 ± 0.08 s^{-1} and 0.36 ± 0.02 s^{-1} and a K_M of 26 ± 3 mM and 3.4 ± 0.9 mM, respectively (Figure3.17 A and B). By comparing the difference in ArgE catalytic activity for substrates **1a, 1b** and **1c**, it is clear that the low affinity of this enzyme towards *N*-Ac-PT is probably caused by the presence of a negatively charged group (phosphinic acid, pKa ~1.0) on the amino acid side chain, which is the main difference between the three structurally analogous substrates (Figure 3.15).

This preliminary screen shows that ArgE has a much wider substrate scope than expected, being able to accept some *N*-acetylated NCAAs. Further studies are required to understand the full potential of this biocatalyst.

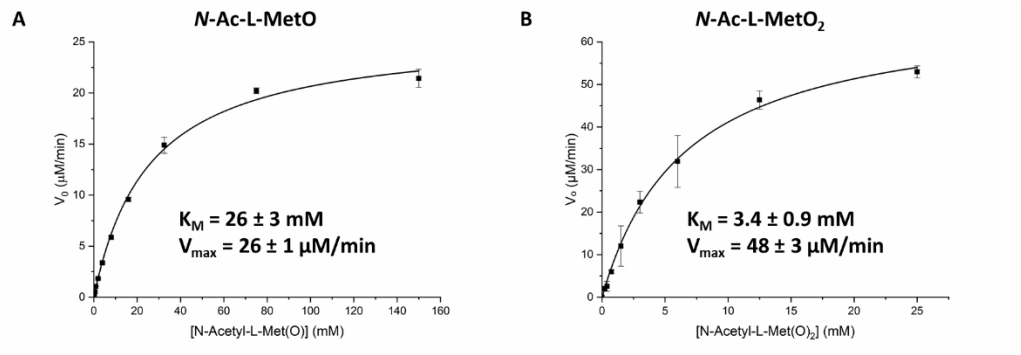


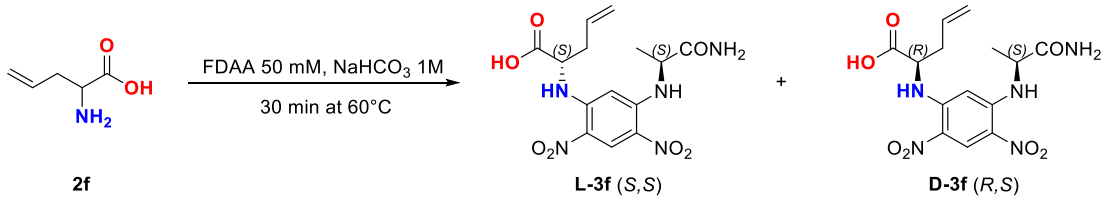
Figure 3.17 Michaelis-Menten plot of ArgE for **A)** *N*-Ac-L-MetO and **B)** *N*-Ac-L-MetO$_2$. Initial rates calculated from the linear portion of the absorbance curves and fitted with the OriginLab software. The reaction mixture contained: ArgE 0.1 mg/mL, $CoCl_2$ 100 µM, L-AAO 70 mU/mL, HRP 10 U/mL and *o*-dianisidine 0.1 mg/mL in NaPi 0.1 M pH 8. The plate was incubated at 40°C for 1 h and absorbance measured at 436 nm every 30 s. Reactions repeated in triplicates, error bars calculated as the standard deviations.

3.8 Development of a quantitative assay for ArgE

3.8.1 Marfey's Reagent Derivatization

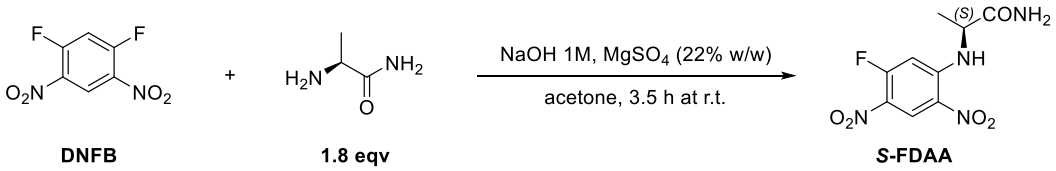
The chiral HPLC assay has been useful to monitor the KRs of *N*-Ac-methionine, however only a few amino acids (as Phe, Tyr and Trp) have a suitable absorbance in the UV-Vis spectrum. A more comprehensive approach is required to efficiently study the biocatalytic potential of ArgE towards non canonical *N*-Ac-AAs.

Various HPLC *pre*-column derivatization methods, such as: ninhydrin[153] *o*-phthalaldehyde (OPA)[154,155] and 4-chloro-7- nitrobenzo-2-oxa-1,3-diazole (NBD),[156] have been developed over the years to quantify amino acids. However, the Marfey's reagent (MR) protocol was particularly attractive, given the chiral nature of the derivatization reagent. Marfey's reagent, or (*S*)-2-((2,4-difluoro-5-nitrophenyl)amino)propenamide (FDAA),[157] possesses a chiral center on its structure, thus it can react with a racemic mixture of amino acids to form two diastereoisomers (Scheme 3.4), which can be easily separated on a C18 column by reverse phase HPLC. This method would be extremely useful, not only to determine the biotransformations conversions, but also to measure the enantiomeric excess (*ee*) of the products. Furthermore, *N*-Ac-AAs will not interfere with the derivatization reaction since FDAA reacts exclusively with free amines. MR derivatives have a strong absorbance at 360 nm and can be easily detected by LC-MS on positive ion mode.[158] They degrade quickly if exposed to direct light, but can be stable in the dark at room temperature for 48 h.



Scheme 3.4 Marfey's reagent (FDAA) derivatization of a standard solution of racemic allylglycine (**2f**).This results in the formation of two diastereoisomers: **L-3f** (*S,S*) and **D-3f** (*R,S*).

FDAA is commercially available but expensive (100 mg for 157 £ by Merck). Since large amounts of the reagents are required, to carry out all the appropriate test to optimize the assay protocol, a batch of MR was prepared in house. FDAA was synthesised following the original protocol reported in 1984 by Marfey[157] (Scheme 3.5) *via* a nucleophilic aromatic substitution of one of one of the fluorine of 1,5-difluoro-2,4-dinitrobenzene (DFNB) by the amino group of (*S*)-2-aminopropanamide in the presence of a strong base. Pure MR crystals were isolated with a yield of 47.7%.



Scheme 3.5 Synthesis of FDAA from DNFB and L-alanine amide.

The derivatization protocol was tested by reacting FDAA with a standard solution of the amino acid allylglycine (**2f)**, using the protocol reported by Wu (Scheme 3.5).[159] The derivatization mixture was filtered and analysed by HPLC. The absorbance chromatogram, recorded at 360 nm, shows two peaks which eluted with a retention time of 23.6 min for **L-3f** (*S,S*) and 26.7 min for **D-3f** (*R,S*) (Figure 3.18).[160] Analysis by LC-TOF MS confirmed the presence of the derivatization products.

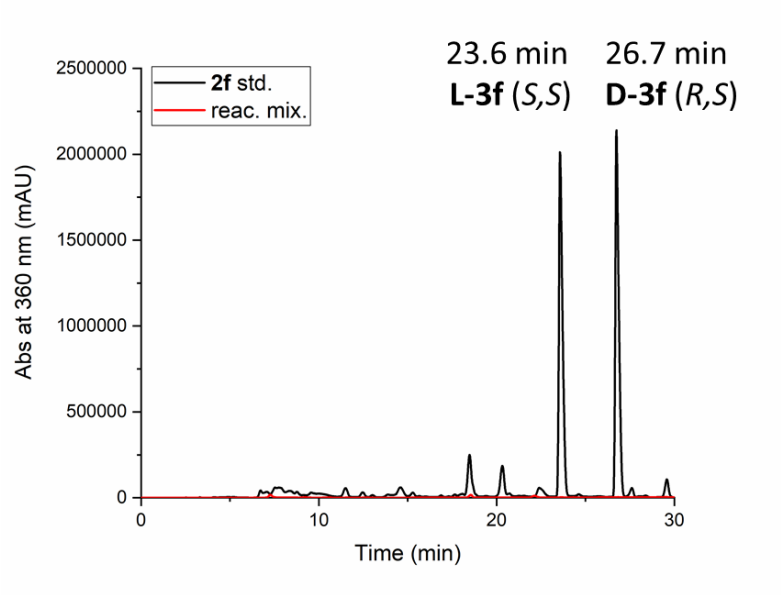
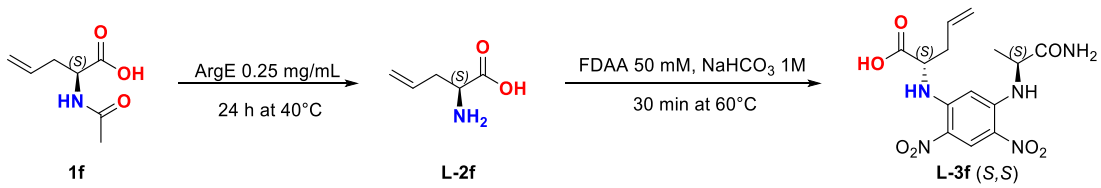


Figure 3.18 HPLC absorbance traces of MR derivatization mixture of a standard L-allylglycine **2f** solution (black) and a reaction mixture (red). The samples eluted on a C18 column with a linear gradient of 15% to 50% ACN 0.1% FA in water 0.1% FA, over 30 minutes. Flow 1.0 mL/min and λ = 360 nm at r.t.

N-Ac-allylglycine (**1f**) was used as substrate for ArgE KR, the reaction was incubated for 24 h at 40°C and product **2f** formation was confirmed by TOF MS analysis. The MR protocol was further tested by reacting FDAA with *N*-Ac-allylglycine KR reaction mixture (Scheme 3.6). The derivatization solution was analysed by HPLC and LC-TOF MS, however no traces of allylglycine-MR derivatives were detected by either technique (Figure 3.18).



Scheme 3.6 *N*-Ac-allylglycine (**1f**) was hydrolysed in a KR with ArgE (0.25 mg/mL) in NaPi buffer 0.1 M pH 8. The KR was incubated 24 h at 40°C; after quenching the biotransformation a sample was derivatised with MR to give as expected product the FDAA derivative **L-3f** (S,S) diastereoisomer.

Attempts to optimise the reaction protocol by changing the incubation temperature, the derivatization time, the reagents concentration, the base and buffer used, did not prove successful.

The assay reproducibility is quite poor, especially when a reaction mixture is used for the derivatization, instead of a standard amino acid solution. To be reliable, the MR

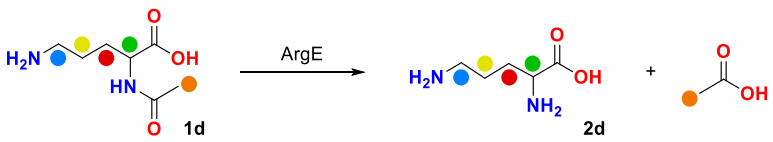
derivatization protocol requires a long optimization. Another more immediate solution was selected to monitor the hydrolysis conversions.[157]

3.8.2 ^{1}H-NMR assay

3.8.2.1 Proof of concept

^{1}H-NMR spectroscopy is a good candidate as a technique to monitor the enzyme catalysed deacetylations. The ^{1}H-NMR spectra of free and *N*-acetylated amino acids have at least two very distinctive signals, which undergo a shift during hydrolysis. The C-α-proton shifts from 4-5 ppm to 3-4 ppm and the acetate methyl group shifts from 2.1-1.9 ppm to 1.9-1.8 ppm following deacetylation. By comparing the integrals of these signals, it should be possible to calculate the relative conversion of the reaction.[161–163]

For proof of concept, a biotransformation was set up, incubating *N*-Ac-L-ornithine (**1d**, 20 mM) with ArgE (0.25 mg/mL) and $CoCl_2$ (100 µM) at 40°C for 24 h in a final volume of 500 µL (Scheme 3.7). The formation of the product **2d** was confirmed by LC-TOF MS analysis. The biotransformation was quenched by addition of HCl conc. to precipitate the enzyme. 100 µL of D_2O was mixed in an NMR tube with 400 µL of reaction mixture. The ^{1}H-NMR spectrum of the sample was recorded on an AVA 500 (500 MHz) NMR, with a solvent signal suppression experiment (16 scans).



Scheme 3.7 Scheme for the ArgE catalysed deacetylation of *N*-Ac-L-Orn (**1d**).The hydrogens of substrate and products are colour coded and reported in: green (α), red/yellow (β/γ), blue (δ) and orange (acetate CH_3 group).

As shown in Figure 3.19, the biotransformations reached completion after 24 h incubation and the C-α-proton and acetate signals can be potentially used to monitor the conversions. However, the α-proton signal of *N*-acetylated amino acids falls close to the residual solvent peak of D_2O (4.79 ppm). Depending on the nature of the amino acid side chains, the two can overlap or the intensity of the signal can potentially be altered during the solvent suppression experiment. The acetate signal falls far from the water peak and can be used to obtain a more accurate estimate of the deacetylation conversion. Furthermore, the acetate peak is formed by three equivalent protons, which should give a better intensity/noise ratio and higher assay sensitivity; indeed, sodium acetate is a common internal standard used for quantitative

NMRs experiments, where the absolute concentration of the species in solution is determined.[164]

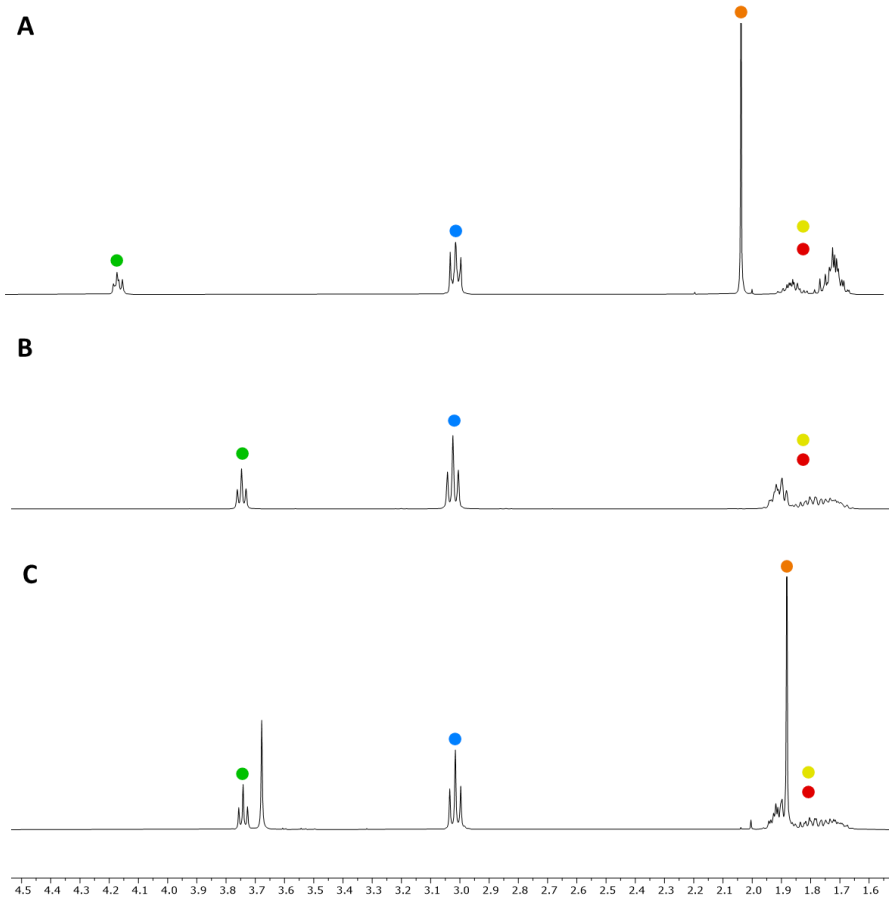
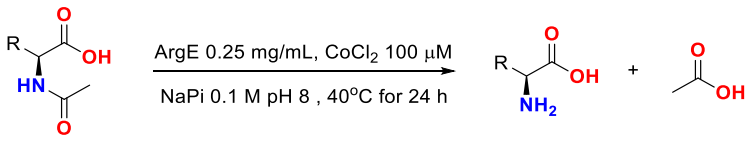


Figure 3.19 ^{1}H-NMR spectrums with colour coded signals of **A)** *N*-Ac-L-Orn (**1d**) standard, **B)** L-Orn (**2d**) standard and **C)** ArgE biotransformation reaction mixture with L-Orn formation (**2d**). The proton signals are reported in: green (α), red/yellow (β/γ), blue (δ) and orange (acetate CH_3 group).

3.8.2.2 Conversion of canonical *N*-Acetyl-L-Amino Acids

Having successfully used *N*-Ac-L-Orn (**1d**) as proof of concept, we then set up nineteen biotransformations using *N*-acetylated L-amino acids 20 mM (with the exception of serine, used as a racemic mixture), ArgE 0.25 mg/mL and $CoCl_2$ (100 µM) and incubated at 40°C for 24h (Scheme 3.8). The reactions were quenched by addition of HCl_{conc} and centrifuged to separate the precipitated enzyme from the supernatant. A ^{1}H-NMR spectrum was recorded for each sample.



Scheme 3.8 Reaction scheme of the ArgE catalysed *N*-acetyl-L-amino acids hydrolysis.

The conversions, reported in Table 3.4, agreed with the kinetic data determined with the L-AAO assay (Table 1.1), with low to no activity detected for aromatic substrates and a preference for hydrophobic or basic groups on the *N*-Ac-AA side chains. In addition, *N*-Ac-L-cysteine and proline were identified as ArgE substrates, while no activity was detected towards *N*-Ac-L-Asp and Glu. These results were expected, since ArgE already showed a low tolerance for *N*-acetylated amino acids with negatively charged groups on the side chain, as PT.

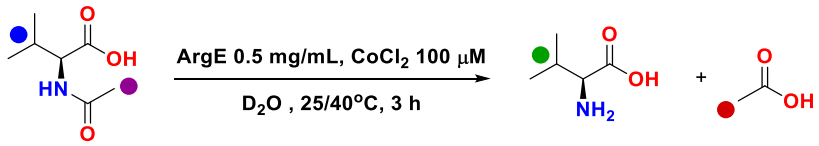
L-AA	Conv (%)	m/z $[M+H]^+$
Ala	93.9	90.0
Ile	98.3	132.1
Leu	99.3	132.1
Pro	53.7	116.0.
Val	85.4	118.1
Phe	2.10	166.1
Tyr	N.D.	N.D.
Trp	N.D.	N.D.
Asp	N.D.	N.D.
Glu	N.D.	N.D.
Arg	99.3	124.1
His	96.1	156.1
Lys	98.8	147.0
Ser	44.9	106.1
Thr	27.3	147.1
Cys	77.8	120.1
Met	100	122.1
Asn	95.4	150.1
Gln	60.6	133.1

Table 3.4 ArgE biotransformation products conversion percentages calculated from the ^{1}H-NMR spectrums and m/z values from TOF-MS analysis on positive ionization modes.

3.8.2.3 Time dependant experiment

The ^{1}H-NMR assay can potentially be exploited to monitor the reaction conversion over time. For the time dependant experiment *N*-Ac-L-Val was chosen as substrate, since this compound

has another signal, the isopropyl group at ~1 ppm, which falls close to the acetate peak and shifts in the spectrum during the hydrolysis.



Scheme 3.9 Reaction scheme of the ArgE catalysed *N*-Ac-L-Val hydrolysis for the time dependant ^{1}H-NMR experiment. The hydrogens of substrate and products are colour coded and reported in: blue and green for the iPr group and purple and red for the acetate CH_3 group.

Two enzymatic reactions were set up at 30 mM substrate concentration in NMR tubes using D_2O as solvent (Scheme 3.9). Since ArgE is a hydrolytic enzyme and its rate limiting step is the water deprotonation, as reported by Blanchard *et. al.,*[141] a higher enzyme concentration (0.5 mg/mL) was employed for these experiments, to compensate for the loss of activity due to the solvent isotopic effect.

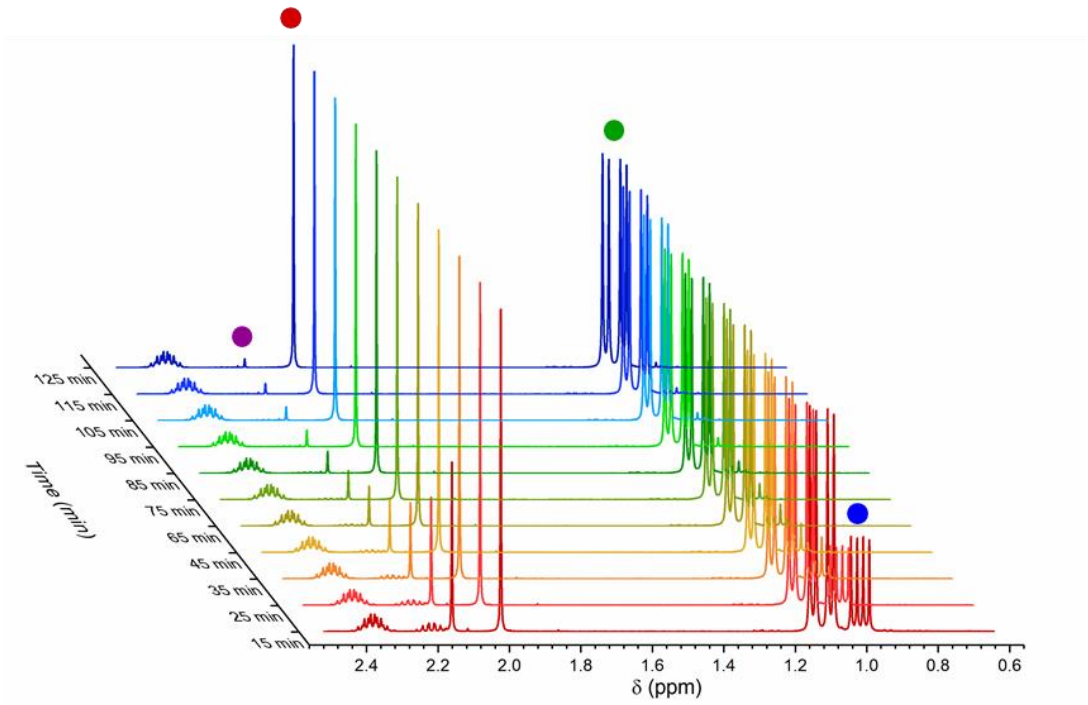


Figure 3.20 Superimposition of ^{1}H-NMR spectra of the *N*-Ac-L-Val hydrolysis at 40°C during incubation times 15 to 125 min. The spectra intensity values were normalised between 0 and 100. The proton signals of the starting material (purple and blue) and the product (red and green) monitored during the reaction, are colour coded as described in Scheme 3.9. Data intensity normalised against the free acetate singlet peak.

The two reactions were incubated at 25°C and 40°C respectively and a ^{1}H-NMR spectrum was recorded on an AVA 400 (400 MHz) NMR every 10 minutes for 2.5 h (Figure 3.20), starting

after 15 min from the addition of ArgE. This lag phase was necessary to set up the experiment and to give enough time to the NMR probe and the reaction to reach the desired temperature

By analysing the superimposed ^{1}H-NMR spectra it is easy to monitor the progress of the reactions with the *N*-acetate peak intensity (2.16 ppm) decreasing over time, while the free acetate signal (2.0 ppm) grows in a time-dependant manner; the isopropyl group signal for starting material and product follow the same pattern (Figure 3.20). Conversions were calculated from the integrals of the free (2.0 ppm) and *N*-bound acetate peaks (2.16 ppm). As expected, given ArgE's optimal reaction temperature, the reaction at 40°C is favoured. It reaches full conversion after 2 h incubation, compared to the 81% conv. obtained for the hydrolysis incubated at room temperature (Figure 3.21).

The reported ^{1}H-NMR assay is reliable and reproducible, plus it does not require the preparation of a calibration curve to obtain quantitative analysis and the samples are easy to prepare. This assay will be employed in all future experiments to measure the conversion of acylase catalysed hydrolysis reactions.

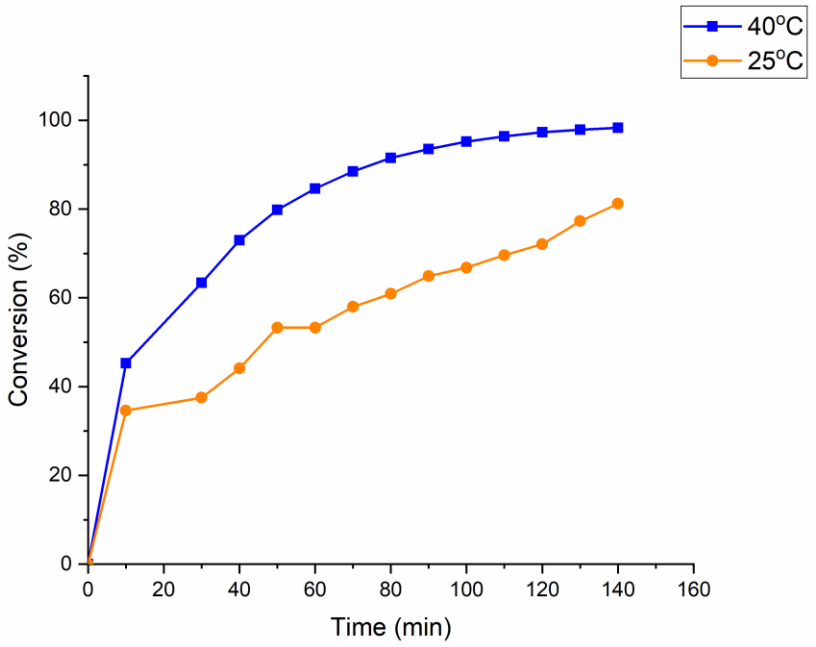


Figure 3.21 Formation of L-Val over time from the ArgE catalysed biotransformations incubated at 40°C (blue) and 25°C (orange). Conversion percentages calculated from the integrals of the acetate signals from the ^{1}H-NMR spectrums.

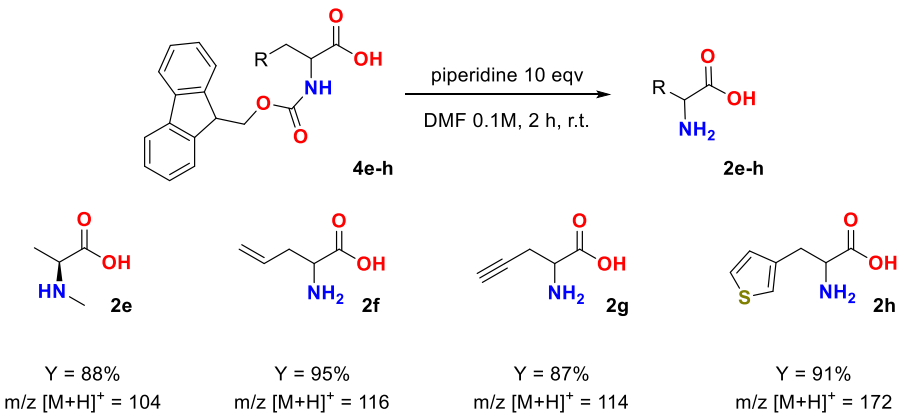
The results obtained employing the ^{1}H-NMR assay are concordant with the data collected on ArgE thermostability and substrate scope, using the spectrophotometric L-AAO assay

(Section 3.3 and 3.6.), proving that ^{1}H NMR spectroscopy can be employed to efficiently monitor enzymatic reactions and to obtain reliable data. Furthermore, the new assay can potentially be applied for the determination of the acylase kinetic parameters, by reducing the 15 min lag phase between the start of the reaction and first spectrum recorded, to obtain a better estimate the initial rate values.

3.9 Expanding ArgE substrate scope

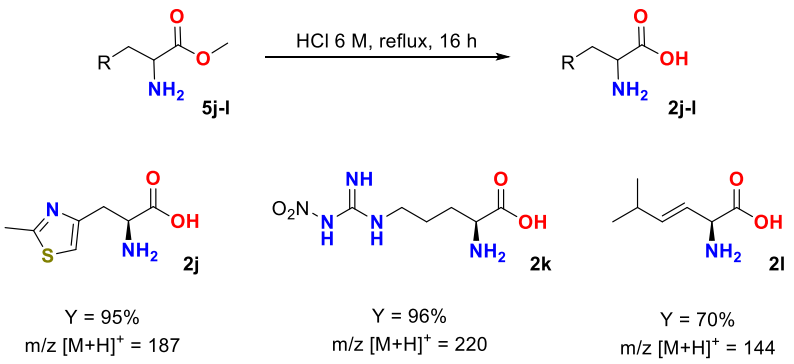
3.9.1 Non canonical *N*-acetyl amino acid library synthesis

The novel ^{1}H-NMR assay was further employed to expand ArgE substrate scope by screening a library of *N*-Ac-NCAAs. The *N*-acetyl amino acids (**1a** and **1e-t**) were prepared from their respective free AAs or derivatives, a gift from Syngenta (Jealott's Hill, UK), with the exception of *N*-acetyl-2-amino-3-(1H-1,2,4-triazol-1-yl)propanoic acid **(1i)**, which was obtained in its acetylated form (gifted by Syngenta, Jealott's Hill, UK). While, *N*-acetyl: L-MetO (**1b**), L-MetO$_2$ (**1c**), L-Orn (**1d**), 4-F-Phe (**1u**), 3-F-Phe (**1v**) and 2-F-Phe (**1w**) were purchased from Sigma-Aldrich.



Scheme 3.10 Deprotection of FMOC-AAs.

The FMOC AAs (**4e-h**) were deprotected using an excess of piperidine in DMF, the reaction was complete after 2 h at room temperature with stirring (Scheme 3.10). While the methyl and ethyl esters (**5j-l**) were hydrolysed overnight under reflux with 6M HCl (Scheme 3.11) to obtain the free amino acids products (**2e-l**).

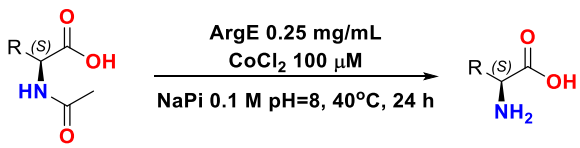


Scheme 3.11 Protocol for ester hydrolysis.

The free amino acids and **2e-u** were acetylated with the same protocol used for **2a** (Section 3.7.1): treatment with acetic anhydride (1.2 eqv) and excess NEt_3 (2.1 eqv), overnight with stirring (Scheme 3.3). It was noted that product **1l** was isolated as a 1:1 mixture of *cis* and *trans* stereoisomers.

3.9.2 ArgE catalysed biotransformations

The first set of reactions was prepared by mixing **1b-q** (20 mM), ArgE (0.25 mg/mL) and $CoCl_2$ (100 µM) in 0.1 M NaPi buffer pH 8. The biotransformations, along with positive and negative controls were incubated for 24 h at 40°C (Scheme 3.12). After quenching the reactions with the addition of HCl conc., the ^{1}H-NMR spectrum of the reaction mixtures was registered, and the reaction conversions calculated (Figure 3.22).



Scheme 3.12 Protocol for the ArgE catalysed hydrolysis of *N*-Ac-L-NCAAs.

ArgE accepted most of the *N*-Ac-NCAAs, with the exception of the sterically hindered substrate: *N*-Ac-*tert*-Leu (**1p**) and *N*-Ac-*tert*-Bu-Ala (**1q**). The enzyme was not able to accept *N*-acetyl methyl-L-alanine (**1e**), however it is active towards other substrates with tertiary amide bonds, such as *N*-Ac-L-Pro.

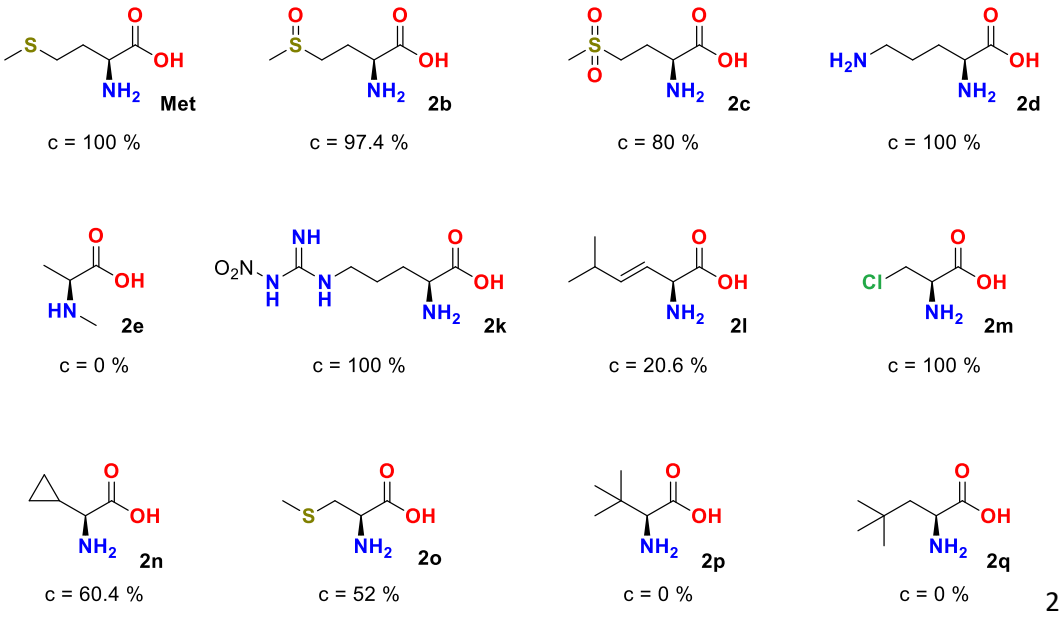


Figure 3.22 ^{1}H-NMR calculated conversions for the enzymatic deacetylation of the *N*-Ac-NCAA library.

One of the lowest conversions (20.6%) was obtained using *N*-acetyl (E)-2-amino-5-methylhex-3-enoic acid (**1l**) as the substrate; this compound is present in solution as a mixture 44:56 of two amide rotamers, *cis* and *trans* (Figure 3.23 A).[165] The presence of a mixture of rotamers was confirmed by conducting a complete NMR charachterization at various temperatures of an **1l** standard sample, which excluded the presence of a mixture of E,Z stereoisomers, see Appendix 22. Instead, rotamers formation due to the presence of the acetamide group is suggested by the coalescence of the ^{1}H-NMR peaks at high temperature (Appendix 22 D). For this compound it was deduced that the *trans* is the major rotamer in solution, since this particular conformation can be stabilised by the formation of an intramolecular H-bond between the amide N-H and the carboxylic acid group, which should be deprotonated in the reaction conditions (buffer pH 8). By analysing the ^{1}H NMR of the biotransformation mixture in the double bond region it was deduced that only the *cis* minor rotamer is accepted by ArgE. As reported in Figure 3.23, the *trans* signal (in blue) remains unaltered after the enzymatic reaction. However, a new peak appears at 5.92 ppm (purple), suggesting that only *cis* rotamer is accepted by ArgE and converted into the free amino acid **(E)-2l**.

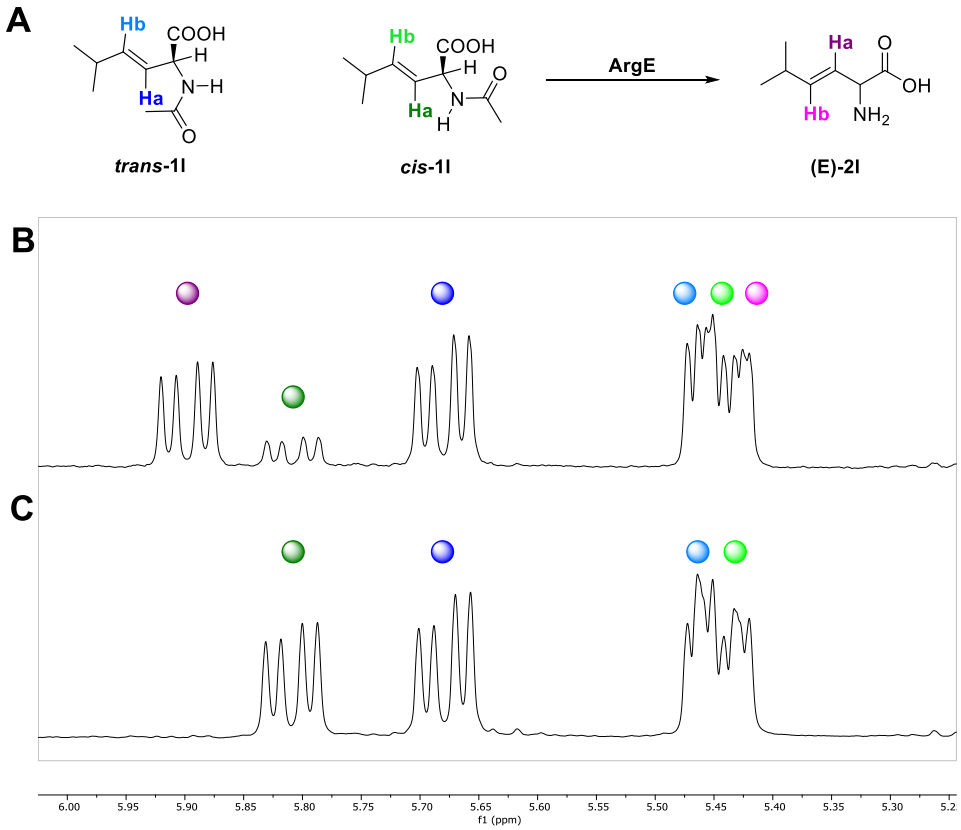


Figure 3.23 A) Reaction scheme for the biocatalytic hydrolysis of (E)-2-acetoamido-5-methylhex-3-enoic acid (**1l**). **B)** Double bond region (5-6 ppm) of the ^{1}H-NMR spectrum of **1l** reaction mixture after 24 h incubation at 40°C with ArgE (0.25 mg/mL) and $CoCl_2$ (100 µM). **C)**. Double bond region (5-6 ppm) of the ^{1}H-NMR spectrum of (E)-2-acetoamido-5-methylhex-3-enoic acid (**1l**). The two double bond protons (**Ha** and **Hb**) are reported in two shades of: blue (the starting material *trans* rotamer), green (the starting material *cis* rotamer) and purple (the product *trans* isomer)

3.9.3 PT and Glu Studies

The results reported in Section 3.7.2, 1.7.3 and 3.8.2.2 suggest the lack of activity towards *N*-Ac-PT **1a** is caused by the presence of the phosphinic acid moiety on the side chain of the *N*-Ac-AA. Indeed, ArgE showed a high activity for two substrates structurally analogous to *N*-Ac-PT, **1b** and **1c**, but it was completely inactive towards other *N*-acetyl amino acids with negatively charged groups in their structure, such as *N*-Ac-Glu and Asp (Figure 3.24).

To definitively prove this hypothesis, the activity of ArgE for three new **1a** analogues: *N*-Ac-L-Glu, *N*-Ac-L-5-methoxy-Glu (**1r**) and *N*-acetyl dimethyl L-glutamate (**1s**), was tested. The same biotransformation protocol described in Section 3.9.2 was employed.

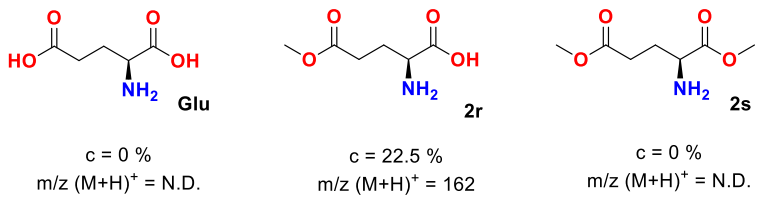
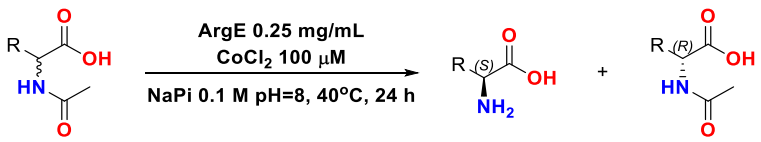


Figure 3.24 Comparison between ArgE activity towards *N*-Ac-L-Glu, *N*-Ac-L-5-methoxy-Glu (**1r**) and *N*-acetyl dimethyl L-glutamate (**1s**). Conversion measured by ^{1}H-NMR spectroscopy and m/z values obtained *via* LC TOF-MS.

No conversion was detected for *N*-Ac-L-Glu and for the di-methyl ester substrate **1s**. Formation of product **2r** (c = 22.5%) was observed by ^{1}H-NMR spectroscopy and LC TOF-MS. By "hiding" the negative charge of glutamate side chain, *via* methyl ester formation (**1r**), the enzyme was able to recognise and bind this substrate, catalysing the formation of the γ-methyl ester product (**2r**) However, this newfound activity is completely lost when both carboxylic groups of glutamate are methylated. These data suggest that an α free carboxylic acid group is required for substrate recognition, while the presence of a negatively charged group on the side chain of the substrate is not tolerated by ArgE.

3.9.4 Kinetic Resolution on *N*-Ac-NCAAs

Up to this point only enantiopure substrates were used for screening, with the exception of **1a**. The second section of the *N*-Ac-NCAAs library is made up by racemic compounds, which were employed to perform an ArgE catalysed kinetic resolution; the protocol reported in Section 1.9.2 was employed to run these reactions (Scheme 3.13).



Scheme 3.13 ArgE catalysed kinetic resolution of racemic *N*-acetyl NCAAs.

As reported in Figure 3.25, ArgE shows great affinity for unsaturated amino acids derivatives (**1f** and **1g**) and is able to accept *N*-acetyl heterocyclic amino acids (**1h-j**) with good to moderate activity. As expected, no product formation was detected using phenylalanine derivatives as substrates (**1u-w**).

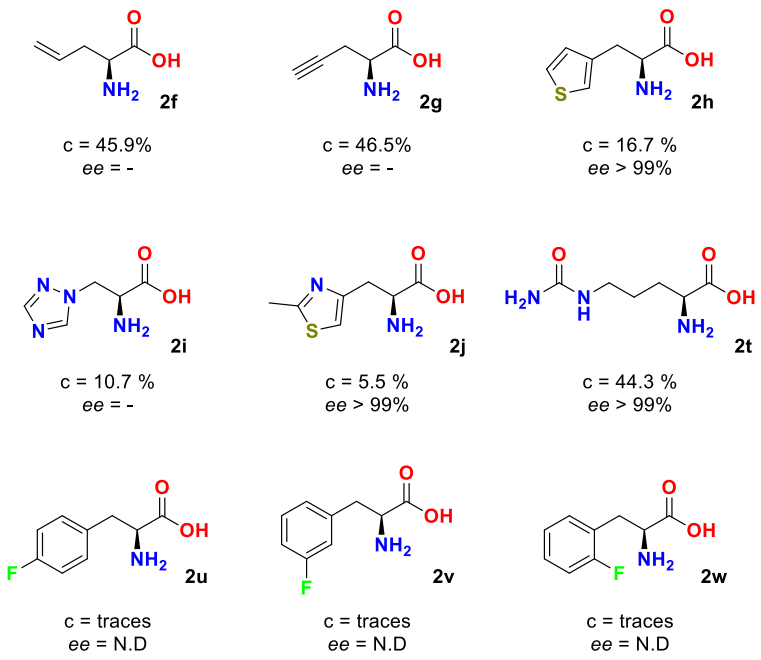


Figure 3.25 ^{1}H-NMR calculated conversions for the enzymatic KR of the racemic *N*-Ac-NCAA.*ee* percentages calculated by chiral HPLC analysis on a Chirobiotic T (Teicoplanin) column, isocratic gradient: elution phase 0.025% TEAA and MeOH (60/40 ratio), flow 0.5 mL/min, T = 40°C and λ = 210 nm. Reactions repeated in duplicates, error bars calculated as standard deviations reported only for triplicates.

The chiral HPLC assay was exploited to estimate ArgE enantioselectivity towards these substrates. Product **2h**, **2j** and **2t** have a weak absorbance at 210 nm (Figure 3.26 A and B), thus it was possible to calculate their enantiomeric excess (> 99%).

The amino acids **2f**, **2g** and **2i** do not absorb in the UV-Vis spectrum, however it was still possible to determine the stereoselectivity of the enzyme by observing the change in their respective *N*-Ac-AA peaks; after 24 h incubation the peak of the L-isomer decrease in intensity, disappearing completely for some samples, while the peak of the D-isomer remains unchanged (Figure 3.26 C and D). This data confirms that ArgE is highly stereoselective.

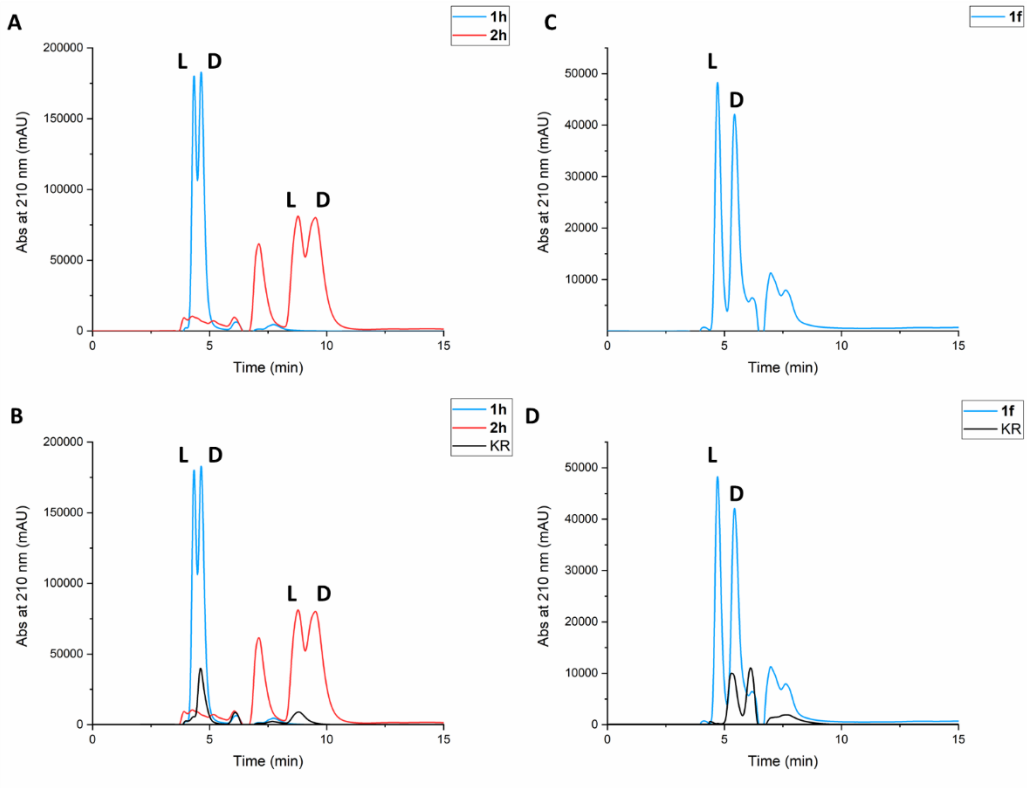


Figure 3.26 Example of chiral HPLC traces employed to study the enantioselectivity of ArgE KR reactions. Comparison between **A) 1h** (blue) and **2h** (red) standards, and **B)** a sample of **1h** kinetic resolution reaction mixture in black. Comparison between **C) 1f** (blue) standard, and **D)** a sample of **1f** kinetic resolution reaction mixture in black. Chiral HPLC analysis on a Chirobiotic T (Teicoplanin) column, isocratic gradient: elution phase 0.025% TEAA and MeOH (60/40 ratio), flow 0.5 mL/min, T = 40°C and λ = 210 nm. Reactions repeated in duplicates, error bars calculated as standard deviations reported only for triplicates.

To improve the reaction conversions, two new kinetic resolution batches were set up with 0.50 mg/mL and 1.00 mg/mL of enzyme, respectively. An increase in the conversion values can be noted for the three heterocyclic substrates: especially **1h** (43%) and **1i** (52%), while no traces formation of **2u**, **2v**, or **2w** was detected by either ^{1}H-NMR spectroscopy or LC-TOF MS, even at the highest enzyme concentration (Figure 3.27).

The best results were obtained using the highest ArgE concentration tested (1 mg/mL); since the conversion values obtained were satisfactory, this enzyme loading will be used for future applications.

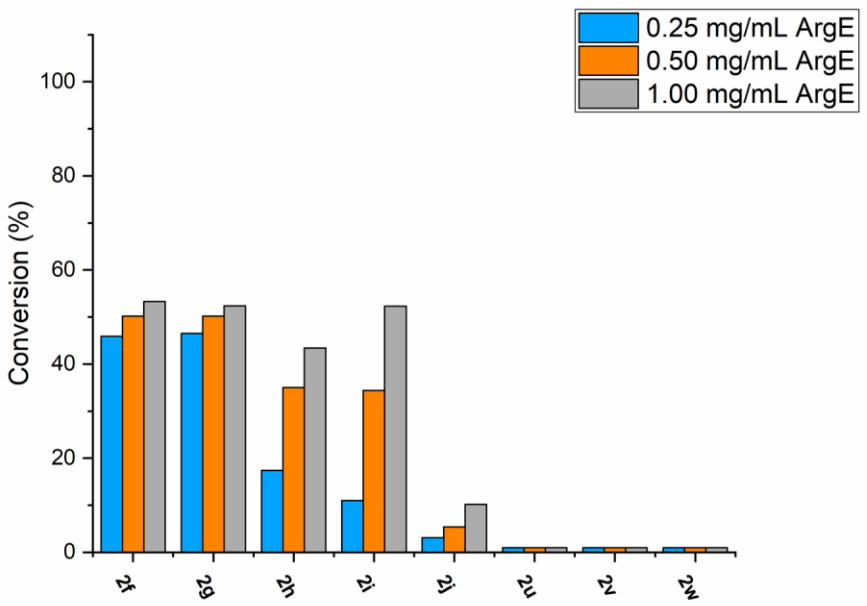


Figure 3.27 Comparison between the conversion values calculated with the ^{1}H-NMR assay for the kinetic resolution of **1f-j** and **1u-w** at increasing enzyme loading: 0.25 mg/mL (blue), 0.50 mg/mL (orange) and 1.00 mg/mL (grey). The reactions contained: substrate (20 mM), ArgE (0.25 mg/mL) and $CoCl_2$ (100 µM) in 0.1 M NaPi buffer pH 8. The KRs were incubated at 40°C, 250 rpm for 24 h. Reactions repeated in duplicates, error bars calculated as standard deviations reported only for triplicates.

3.9.5 Determination of kinetic parameters of ArgE

The NCAAs library was screened for activity with the commercial L-AAO. With the exception of L-*tert*-leucine (**2p**), the oxidase was active for all NCAAs (Figure 3.28). However, this lack of activity of the L-AAO did not cause any drawbacks for the determination of ArgE kinetic parameters. As reported in Figure 3.22 and Figure 3.25, no conversion was measured for ArgE towards the acetylated derivative of L-*tert*-leucine and other steric hindered amino acids.

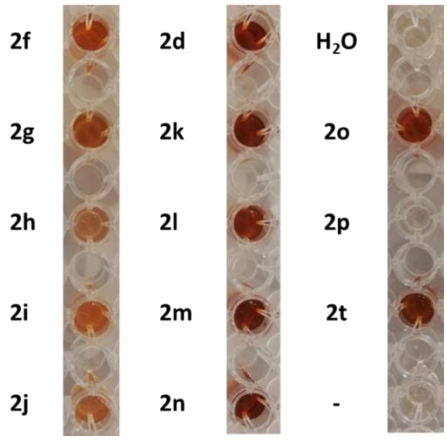


Figure 3.28 Screening of L-AAO activity towards the non-canonical amino acid library. Water used as control. The assay mixture contained: AA 10 mM, $CoCl_2$ 100 µM, L-AAO 100 mU/mL, HRP 10 U/mL

and *o*-dianisidine 0.1 mg/mL in NaPi buffer 0.1 M pH 8. The plate was incubated at 40°C for 1 h and absorbance measured at 436 nm every 30 s. Structures of the amino acids are reported in Figure 3.22 and Figure 3.25.

The kinetic parameters of ArgE for the *N*-acetyl NCAAs library were determined with the continuous L-AAO assay (Table 3.5); it must be noted that ArgE shows a higher activity for *N*-Ac-L-Met (k_{cat} = 4.58 s^{-1}), compared to *N*-Ac-L-Orn (**1d**), its natural substrate. Furthermore, a TON of 0.15 s^{-1} was calculated for substrate **1j**, 5-fold higher than the k_{cat} obtained for compound **1i** ($3 \cdot 10^{-2}$ s^{-1}); however, a 52% conversion was reported for **2i** formation, while the maximum conversion achieved for **2j** at the highest enzyme loading was 10%. ArgE activity may be inhibited by the presence of either substrate (**1j**) or product (**2j**) in high concentration, which could explain this gap registered between TON and reaction conversions. Further tests are required to confirm this theory.

***N*-Ac:**	**K_M (mM)**	**k_{cat} (s^{-1})**	**k_{cat}/K_M ($M^{-1}s^{-1}$)**
1d	3.0 ± 0.3	0.106 ± 0.008	35.250 ± 0.008
1f	2.4 ± 0.2	0.34 ± 0.02	267.870 ± 0.003
1g	1.29 ± 0.08	0.12 ± 0.01	6.47 ± 0.03
1h	-	-	-
1i	36 ± 3	0.0298 ± 0.0008	0.836 ± 0.002
1j	10.03 ± 0.07	0.147 ± 0.002	14.6276 ± 0.0001
1k	2.7 ± 0.1	0.170 ± 0.002	64.3415 ± 0.0005
1l	2.7 ± 0.1	0.170 ± 0.002	64.3415 ± 0.0005
1m	4.1 ± 0.3	0.152 ± 0.009	37.006 ± 0.005
1n	6.2 ± 0.5	0.28 ± 0.02	44.688 ± 0.004
1o	0.4 ± 0.1	0.26 ± 0.01	676.84 ± 0.01
1r	5 ± 1	0.44 ± 0.03	88.78 ± 0.02
1t	12.5 ± 0.4	0.311 ± 0.007	24.8651 ± 0.0008

Table 3.5 Kinetic parameters of ArgE for *N*-Ac-NCAAs. The reaction mixture contained: various substrate concentrations, $CoCl_2$ 100 µM, L-AAO 100-300 mU/mL, HRP 10 U/mL and *o*-dianisidine 0.1 mg/mL NaPi buffer 0.1 M pH 8. The plate was incubated at 40°C for 1 h and absorbance measured at 436 nm every 30 s. Reactions repeated in triplicates.

3.10 Homology model and Mechanistic Studies

3.10.1 Homology Model

To complement the activity and substrate screenings performed, a structural study was conducted. Since the ArgE crystal structure has not been resolved, a homology model was built with the free access RaptorX software (Figure 3.29)[166] using the enzyme amino acid sequence. Five homologues with a resolved X-ray structure were used as templates: DapE

from *Haemophilus. influenzae* (31.5 % sequence identity, PDB entry code 3IC1), DapE from *Neisseria meningitis* (29.0 % sequence identity, PDB entry code 1VGY), carboxypeptidase G2 from *Pseudomonas* sp. (30.3 % sequence identity, PDB entry code 1CG2), *N*-acetyl-ornithine deacetylase from *Rhodopseudomonas palustris* (27.5 % sequence identity, PDB entry code 3PFO) and *N*-acetyl-ornithine deacetylase from *Spaerobacter thermophilus* (29.0 % sequence identity, PDB entry code 4Q7A). These proteins, like ArgE, are members of the M20 peptidase superfamily and are di-zinc dependant hydrolases. Furthermore, they are homodimers and each monomer contains a catalytic and a dimerization domain.[167]

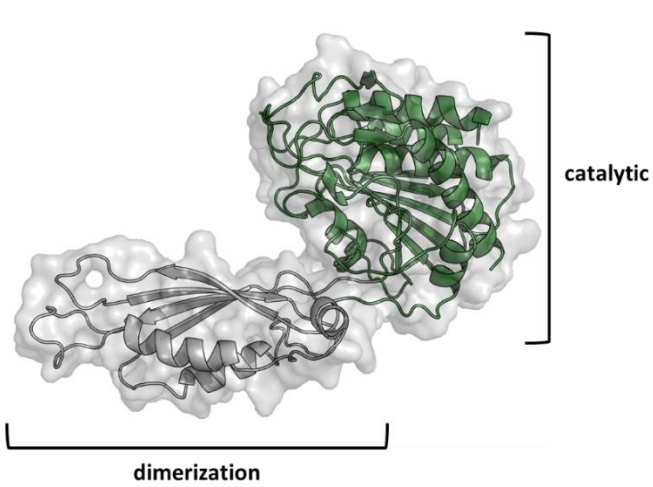


Figure 3.29 *E. coli* ArgE homology model built with the free access software RaptorX, based on the crystal structure of: DapE from *H. influenzae* (3IC1), DapE from *N. meningitis* (1VGY), carboxypeptidase G2 from *Pseudomonas sp.* (1CG2), *N*-acetyl-ornithine deacetylase from *R. palustris* (3PFO) and *N*-acetyl-ornithine deacetylase from *S. thermophilus* (4Q7A). Catalytic domain in green and dimerization domain in silver.

In the catalytic domain each metal ion is coordinated to a glutamate and a histidine residue, with a bridging aspartic acid and water molecule completing the penta-coordination. These residues are involved in maintaining the enzyme's structural integrity but they also play a role during catalysis; additionally, another residue (Glu or Asp) acts as general acid/base, deprotonating the water and starting the reaction.[167] Sequence alignment of ArgE with its five homologues (Chapter 8.6) shows that the residues H80, D112, E144, E145, E169 and H355 (H67, D100, E134, E135, E163 and H349 for *H. influenzae* DapE) are highly conserved in the M20 peptidase family (Figure 3.30).[141]

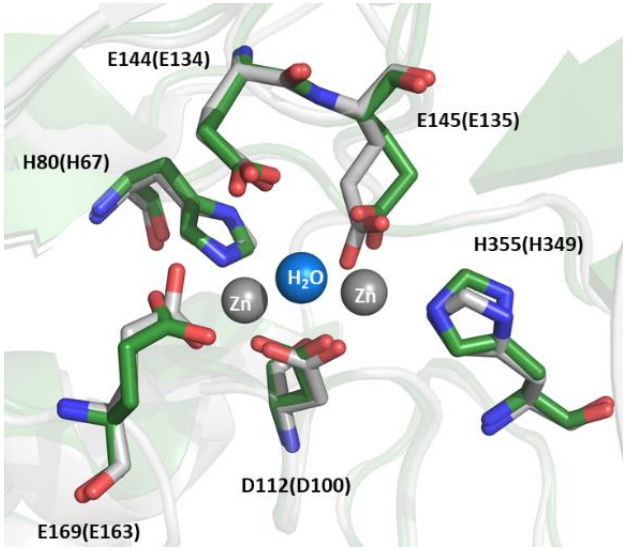


Figure 3.30 Superimposition between the active site of *E. coli* ArgE (in green) and its closest homologue: *H. influenzae* DapE (silver) in complex with two Zn(II) ions (grey spheres); the bridging water is represented as a blue sphere.

The role of residues H80 and H355 was investigated by Holz *et. al.*[168] These two His residues were mutated to Ala. While the H80A mutant was completely insoluble the H355A mutant was successfully expressed and purified, however the mutation caused a significant loss in ArgE activity (measured as the decrease of L-Orn production over time *per* milligram of enzyme) and an increase in the second k_d. This data suggest that H355 is involved in catalysis and occupies the second (weaker) metal binding site.

The role of the three highly conserved glutamate residues has yet to be clarified. As shown in Figure 3.30, E145 and E169 are in proximity of the two zinc ions and suggests that they are involved in the metal binding. However, E144 is quite distant from both Zn^{2+} and it probably acts as a general acid/base, deprotonating the water to start the reaction.

The sequence alignment also revealed the presence of another highly conserved residue in the family. H195, which is located on the dimerization domain, far from the catalytic site. The role of this residue during catalysis was first suggested by Holz *et. al.*[169] by resolving the crystal structure of the *H. influenzae* DapE enzyme in complex with the reaction products: succinic acid (**SNA**) and L,L-di-amino pimelic acid (**L,L-DAP**), PDB entry code 5VO3. The *H. influenzae* DapE is a di-zinc metalloprotein which shares 31% sequence homology with ArgE and was used as a template to build its homology model.

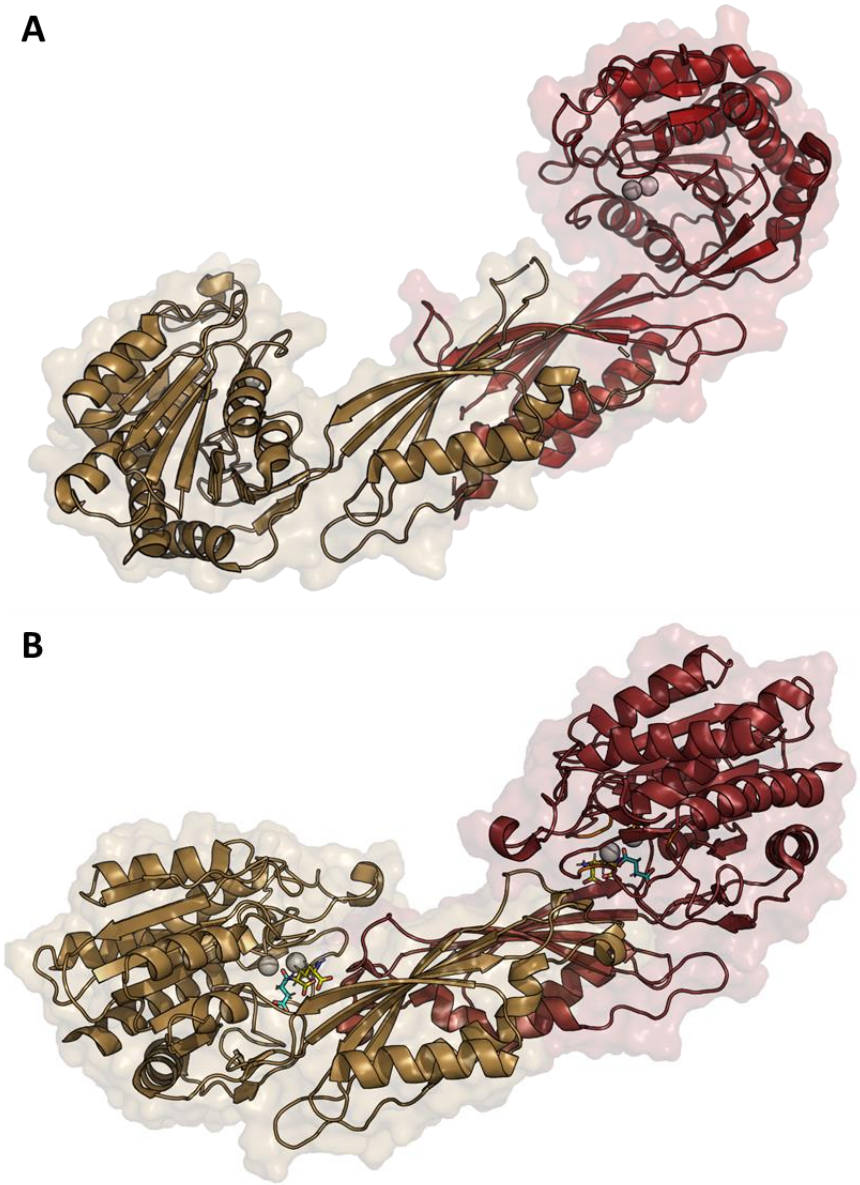


Figure 3.31 Comparison between **A)** open (3IC1) and **B)** close (5VO3) conformation of *H. influenzae* DapE crystal homodimeric structures (chain A in red, chain B in gold, L,L-DAP in yellow, SNA in cyan and Zn(II) ions in grey).

Upon comparing the DapE free (PDB entry code 3IC1,) and products L,L-diaminopimelic acid (L,L-DAP)/succinic acid (SNA)-bound structures (Figure 3.31), a significant conformational change is noted with the catalytic domain twisting and flexing (50° rotation and 10 Å shift)[169] to pass from an open to a close conformation during substrate binding. Holz suggested that this conformational change is caused by the formation of two new interactions between one of the succinic carboxylate groups with Zn(II) and H194 (H195 for ArgE), which would bring the dimerization domain of chain B, where H194 is located, close to the catalytic domain of chain A (Figure 3.32). Another interaction involved in substrate binding stabilization, is

formed between the R258 (R264 for ArgE) guanidinium group and the alpha carboxylate of the substrate (Figure 3.32).

To verify if ArgE follows a reaction mechanism and substrate binding/stabilization mode analogous to *H. influenzae* DapE (31% sequence homology), single point mutants of D112, E144, E145, E169, H195 and R264 were prepared and their activity tested using ArgE WT best substrate: *N*-Ac-L-Met.

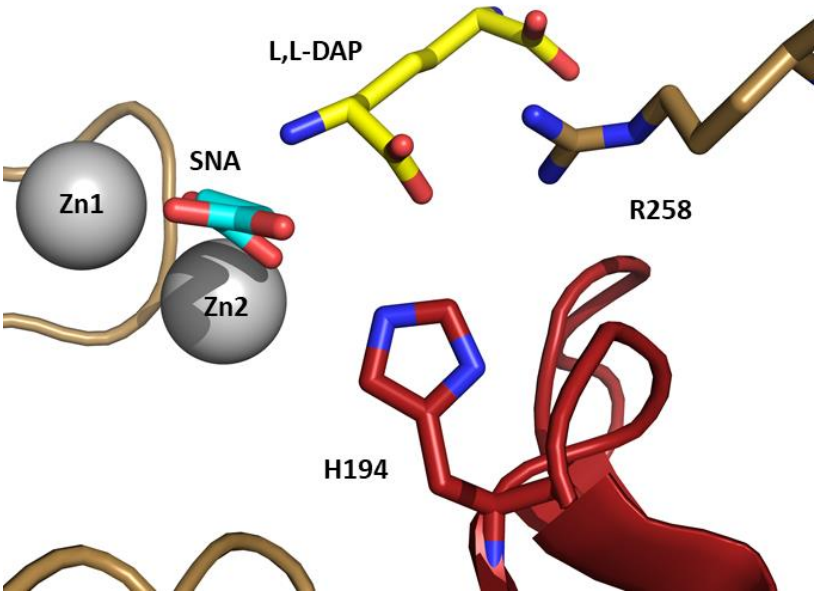


Figure 3.32 *H. influenzae* DapE dimeric structure (chain A in gold and chain B in red) in complex with two Zn(II) ions (grey spheres), succinic acid **SNA** (cyan) and L,L-diaminopimelic acid **L,L-DAP** (yellow). Residues H194:B and R258:A involved in substrate binding and stabilization by H-bonds formation.

3.10.2 ArgE E144A, E145A and E169A Characterization

Mutagenesis primers were designed for the mutants E144A, E145A and E169A, using the method described by Lin and Naismith.[170] The pETHISTEV-ArgE vector was used as a template for site directed mutagenesis (SDM) and the whole plasmid was amplified with the primers incorporating the mutations into the gene. The presence of the desired mutation was confirmed by sequence analysis.

ArgE E144A, E145A and E169A were expressed using the standard ArgE protocol (Chapter 6.3.4), however SDS-PAGE analysis of ArgE E145A and E169A expression test samples, revealed that the mutants were completely insoluble (Figure 3.33 A). ArgE E144A was successfully expressed and purified with a final yield of 10 mg *per* L of culture; the SEC

absorbance chromatogram shows a single peak at 72 mL (Figure 3.33 B), indicating the presence of a homodimeric protein in solution.

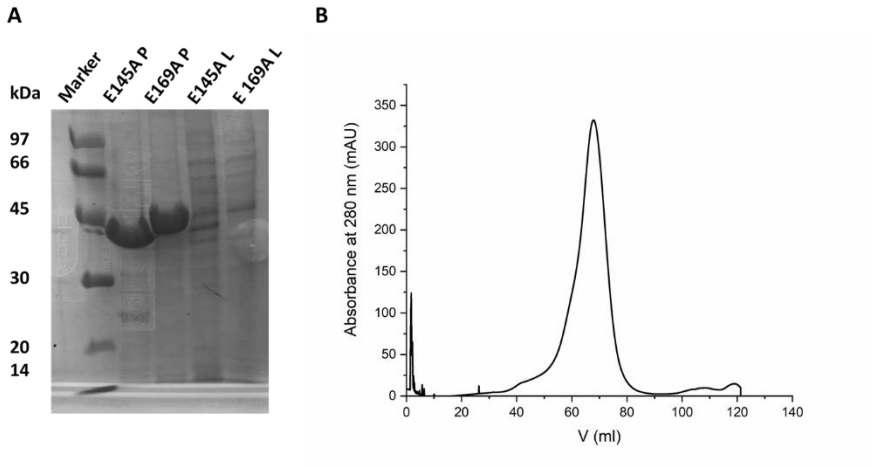


Figure 3.33 Expression and purification of ArgE E144A, E145A and E169A mutants. **A)** 12% SDS-PAGE gel of ArgE E145A and E169A expression tests: lane 1 LMWM (M), lane 2 E145A pellet sample (P), E169A pellet sample (P), lane 4 E145A soluble fraction (L) and lane 5 E169A soluble fraction (L). **B)** SEC chromatogram for ArgE E144A purification on a S200 SEC column.

The SDS-PAGE gel shows a strong band at 45kDa (Figure 3.34 A) and ICP-MS analysis confirm the presence of one zinc bound *per* unit of enzyme. To further verify the presence of the mutation, a sample of purified protein was analysed using denaturing LC ESI-MS. The observed MW of 45312.77 ± 2.08 Da (Figure 3.34 B) was in accordance with the theoretical MW (45311.47 Da) of the enzyme minus the initial methionine, calculated from the amino acid sequence.

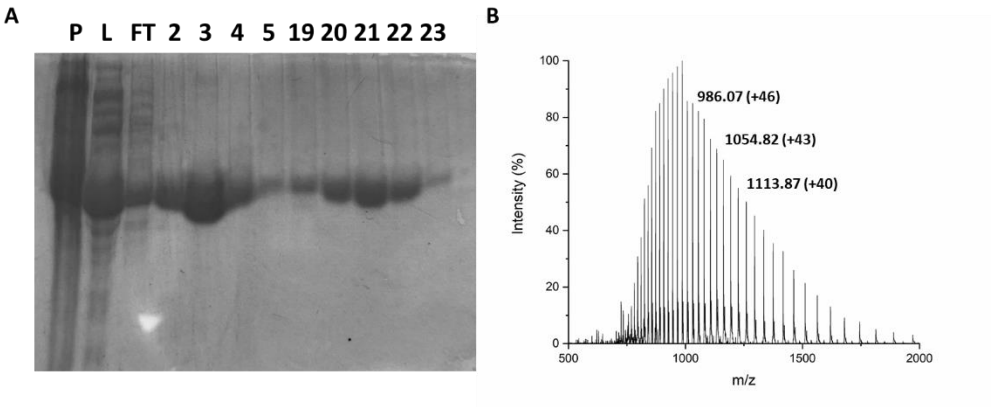


Figure 3.34 Characterization of purified ArgE E144A. **A)** 12% SDS-PAGE gel of ArgE E144A purification: lane 1 LMWM (M), lane 2 CFE (L), lane 3 HisTrap flow through (FT), lane 4-7 HisTrap fractions (3 mL each, enzyme elution start at ~200 mM imidazole) and lane 8-12 S200 fractions (3 mL each). **B)** Denaturating LC ESI-MS of ArgE E144A (20 µM) with a mass of 45312.77 ± 2.08 Da, in

accordance with the MW calculated from the sequence (45311.47 Da). The values reported are m/z with the charges in brackets.

ArgE E144A kinetic parameters for *N*-Ac-L-Met as substrate were obtained employing the high-throughput L-AAO assay (Figure 3.35 A). The mutant was almost completely inactive (Figure 3.35 B); this loss in activity is caused by a significant decrease of the k_{cat} (0.019 ± 0.002 s^{-1}), 250-fold lower than the value calculated for the WT (4.58 ± 0.07 s^{-1}), while the mutant is still able to efficiently bind the substrate, since the K_M, suffered just a slight decrease from 2.4 ± 0.2 mM (WT) to 1.3 ± 0.7 mM (E144A).

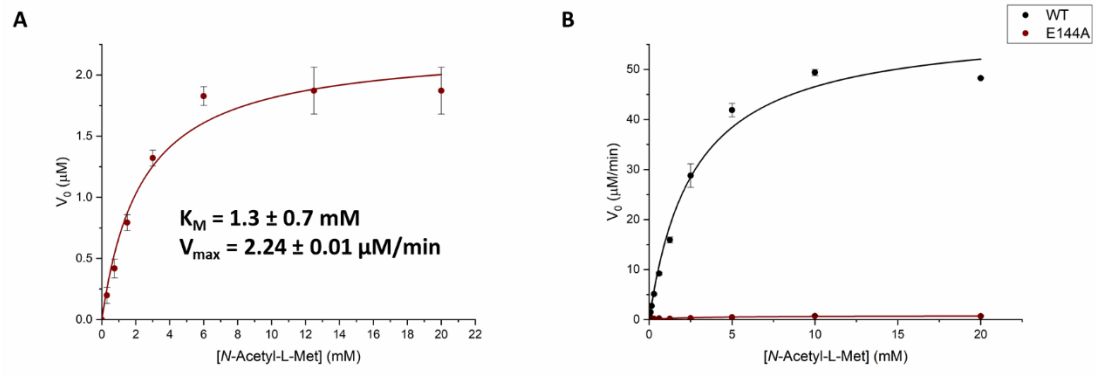


Figure 3.35 Kinetic analysis of ArgE E144A *via* the HT L-AAO assay. **A)** Michaelis-Menten plot of ArgE E144A for *N*-Ac-L-Met. Initial rates calculated from the linear portion of the absorbance curves and fitted with the OriginLab software. The reaction mixture contained: various substrate concentrations, ArgE 10 µg/mL, $CoCl_2$ 100 µM, L-AAO 70 mU/mL, HRP 10 U/mL and *o*-dianisidine 0.1 mg/mL in NaPi 0.1 M pH 8. The plate was incubated at 40°C for 1 h and absorbance measured at 436 nm every 30 s. **B)** Comparison between the Michaelis-Menten curve of ArgE WT (black) and E144A (red) obtained for *N*-Ac-L-Met. Reactions repeated in triplicates, error bars calculated as the standard deviations.

The data collected suggest that the E144 residue assumes a catalytic role in ArgE, acting as a general acid/base catalyst during the water deprotonation step. While the E145 and E169 residues are not directly involved in catalysis, however the insolubility of the two alanine mutants demonstrate the importance of these two residues for cofactor binding and for maintaining the structural integrity of the catalytic site.

3.10.3 ArgE D112A, H195I and R264A Characterization

Mutagenesis primers were designed for the mutants D112A, H195I and R264A, using the method described by Lin and Naismith.[170] The pETHISTEV-*arge* vector was used as template for SDM and the whole plasmid was amplified with the primers incorporating the mutations into the gene. The presence of the desired mutation was confirmed by sequence analysis.

ArgE D112A, H195I and R264A were expressed using the standard ArgE protocol. The three mutants were purified *via* metal affinity chromatography and the fractions containing the enzymes were collected and dialysed overnight at 4°C against storage buffer (50 mM *tris*(hydroxymethyl)aminomethane TRIS pH 8 and 150 mM NaCl). SDS-Page and denaturating LC-ESI MS analysis were carried out to confirm the presence of the desired mutations. Additionally ICP-MS analysis was performed to quantify the zinc concentration bonded to the mutants; the results of the analysis are reported in Table 3.6.

Mutant	Yield (mg/L culture)	$MW_{obs.}$ (Da)	$MW_{calc.}$ (Da)	Zn^{2+} (eqv)
D112A	3.9	45325.07 ± 1.03	45325.49	1
H195I	12	45316.00 ± 2.75	45319.44	1
R264A	13	45281.21 ± 1.52	45284.40	1

Table 3.6 Yield, MW calculated from amino acid sequence, MW observed by denaturing LC-MS and Zn(II) content determined *via* ICP-MS for ArgE D112A, H195I and R264A mutants.

The activity of the mutants towards *N*-Ac-L-Met was explored employing the continuous L-AAO assay (Table 3.7); the Michaelis Menten curve of ArgE R264A did not reach saturation even at high substrate concentration, however the curvature was sufficient to estimate the kinetic parameters of the mutant.

Mutant	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
WT	2.4 ± 0.2	4.58 ± 0.07	1759.653 ± 0.001
D112A	5.0 ± 0.2	0.353 ± 0.002	70.293 ± 0.002
H195I	10 ± 1	0.384 ± 0.004	7.58 ± 0.01
R264A	59 ± 16	0.11 ± 0.01	1.91 ± 0.04

Table 3.7 Comparison between the kinetic parameters obtained for ArgE WT, D112A, H195I and R264A with the continuous L-AAO assay using *N*-acetyl-L-methionine as substrate.

As shown in Table 3.6, the enzyme's zinc and substrate binding were not disrupted by the mutation of the bridge ligand D112; however, the TON of D112A (0.353 ± 0.002 s^{-1}) is 13-folds lower compared to ArgE WT (4.58 ± 0.07 s^{-1}). In the absence of the bridging ligand, the Zn(II) ion is possibly forced to assume a distorted tetracoordinated conformation, similar to those assumed by the cofactor in zinc metalloproteins with a catalytic active site;[171] this change in the geometry may explain the loss in activity detected.

A 13-fold decrease in the k_{cat} was also calculated for the H195I mutant (0.384 ± 0.004 s^{-1}), along with a 5-fold increase in the K_M (10 ± 1 mM). While the kinetic parameters obtained for ArgE R264A shows a TON 40-times lower (0.11 ± 0.01 s^{-1}) and a K_M 25-times (59 ± 16 mM)

higher compared to the WT. The data collected suggest that these two residues are important for the enzyme activity and do play a role in substrate binding; in accordance with the proposed role for the DapE H194 and R258 residues, based on the product-bound structure of *H. influenzae* DapE; suggesting that *E. coli* ArgE and potentially other members of the M20 peptidase family, not only follow the same reaction mechanism, but undergo an analogue conformational change upon substrate binding.

3.10.4 Proposed Mechanism

The data collected from the SDM studies, along with the results previously reported in literature[141,145,168] suggest that ArgE presents a typical co-catalytic binuclear zinc binding site[142,172] and follow a hydroxide mechanism,[167,169,173] as other members of the members of the M20 peptidase family.

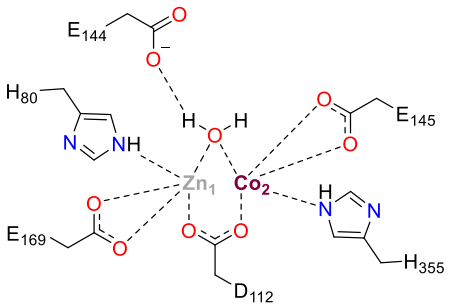
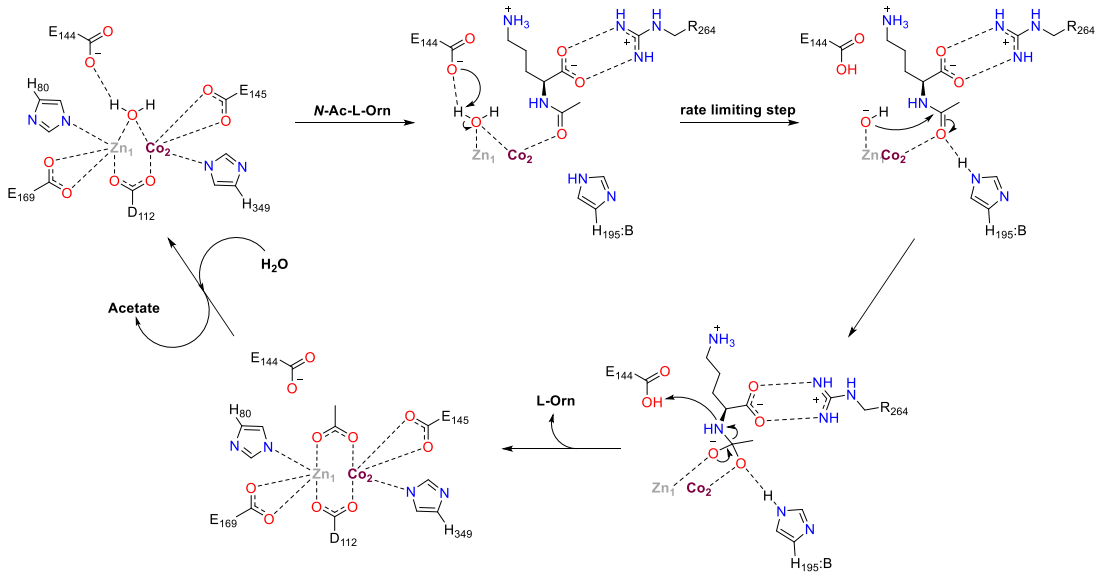


Figure 3.36 Schematic representation of ArgE active site with the Zn^{2+} (grey) and Co^{2+} (purple) bound and E144 ready for catalysis upon substrate binding.

It was noted that after purification ArgE contains only a zinc bound to the "strong" binding site (higher affinity) formed by the E169 and H80 residues; the data reported by Holz[145,168] and the results of the metal binding studies conducted in Section 3.5 suggests that after addition of $CoCl_2$ to the reaction mixture the enzyme will bind a Co^{2+} ion, which is going to occupy the weak binding site, with E145 and H355 acting as ligands (Figure 3.36).

As reported in Scheme 3.14, the reaction starts with substrate recognition and binding; the enzyme undergoes a conformational change, passing from an "open" to a "closed" active conformation. It is probable that, similarly to *H. influenzae* DapE, various ArgE residues, which are going to form the substrate binding site, are involved in this step. The results of the SDM studies suggest that R264 from chain A of the dimer and H195 from chain B (R258·A and H194·B for DapE) are involved in the *N*-Ac-amino acid stabilization with the guanidinium group of R264 forming an hydrogen bond with the α-carboxylate of the substrate, while the acetate C=O, coordinated with the Co^{2+} ion, interacts *via* a H-bond with the imidazole of H195

from chain B, forming an oxyanion hole and increasing the electrophilicity of the carboxylate. The hydroxide is formed from the deprotonation of the bridging water by E144 in the rate limiting step of the reaction;[141] the Zn^{2+} ion acts as a Lewis acid by decreasing the pKa of the water and stabilizing the hydroxide ion, which attacks the electrophilic carboxylate of the acetate. The tetrahedral intermediate, stabilized by coordination to the Zn(II) and Co(II) ions and by a H-bond with H195, collapses thanks to a proton transfer delivered from E144. The amino acid product is released with the enzyme returning to an "open" conformation and a new water molecule substitutes the acetate as bridging ligand reforming the catalytic active site.



Scheme 3.14 Proposed ArgE hydrolysis mechanism of *N*-acetyl-L-ornithine, adapted from the *H. influenzae* reaction mechanism described by Holz *et. al*.[169]

In the absence of $CoCl_2$ the enzyme still retains 70% of its activity (Section 3.5.2). The second metal ion is not fundamental for this class of enzymes,[167,169] since the reaction mechanism can proceed in a similar fashion with the carboxylate group of the acetate being stabilised by hydrogen bonding with H355 and H195.

3.11 Crystal Trials

The homology model built and the SDM studies conducted clarified the role of the active site residues involved in metal binding, water deprotonation and intermediate stabilization. However, to fully comprehend ArgE's substrate selectivity and identify the residues involved

in the amino acid side chain recognition, a substrate bound crystal structure is required. Various ArgE constructs were prepared and tested for the crystal trials.

3.11.1 N-HisTag ArgE WT

N-HisTag ArgE, purified as described in Chapter 6.4, at a concentration of 10 mg/mL was employed to screen 672 conditions with the crystallization kits: MIDAS, STRUCTURE 1+2, PEG ION, Index, PACT, CSS1+2 and JCSG+; all crystallographic studies were conducted in York in collaboration with Prof. Grogan. Unfortunately, only plate shaped salt crystals were obtained from this initial screening.

3.11.2 C-HisTag ArgE WT

The stop codon of the pET28a-ArgE construct was removed by SDM and the *arge* gene was subcloned into a pEBSRCTEVC10HIS plasmid (a gift from Prof. Naismith) using the NcoI and XhoI restriction sites, to obtain a construct with a TEV cleavable 10xHisTag at the C terminus of the protein. pEBSRCTEVC10HIS-ArgE was expressed using the standard ArgE protocol (Chapter 6.3.4), however SDS-page analysis revealed that the protein did not express in these conditions.

The presence of the long 10xhistidine tag at the C-terminus is not tolerated by ArgE; no further attempts were made to improve the construct expression yield.

3.11.3 Non-Tagged ArgE WT

The pETHISTEV-ArgE construct possess a TEV cleavable hexa-histidine tag; TEV protease was expressed and purified as reported in Chapter 6.4 and employed to obtain pure non-tagged ArgE with a four steps purification protocol. The N-His tagged ArgE cell free extract was purified *via* nickel affinity chromatography and the fractions containing the enzyme were collected and incubated with 2 mg/mL of TEV for 16 h and dialysed against ArgE storage buffer (TRIS 50 mM pH 8 and 150 mM NaCl). The untagged protein was separated from TEV by a second nickel affinity chromatography step, the flow-through containing ArgE was collected, concentrated and injected on a pre-calibrated Superdex S200 column. A single peak eluted at 70 mL (Figure 3.37 A), indicative of the presence of a homodimer in solution. ICP-MS and SDS-PAGE analysis (Figure 3.37 B) were performed, confirming that a protein with a MW of ~42 kDa eluted with only 1 eqv of Zn^{2+} bound *per* monomer, indeed it is possible to clearly observe the ArgE MW loss after overnight dialysis with TEV, with the enzyme band

shifting from 45 kDa to 42 kDa on the gel. An observed MW of 42533.08 ± 1.42 Da was obtained by denaturating LC-MS analysis (Figure 3.38 A), matching the predicted MW of 42532.53 Da calculated from the amino acid sequence.

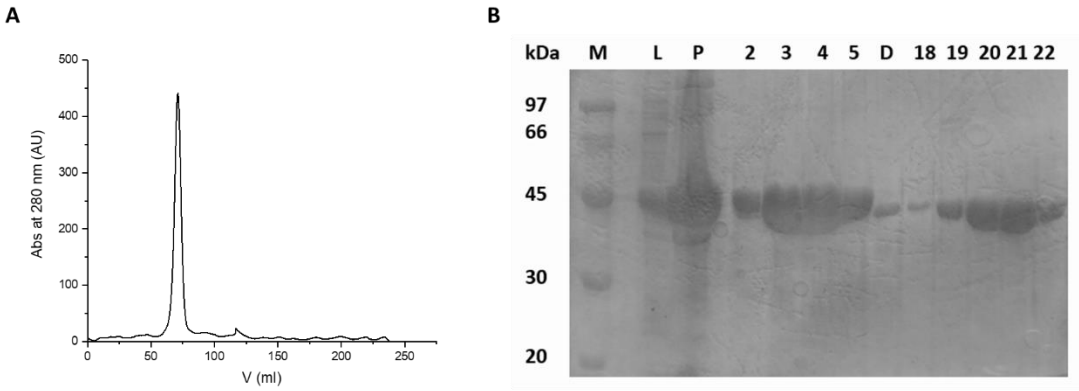


Figure 3.37 Purification of untagged ArgE, following N-His tag cleavage. **A)** SEC chromatogram for untagged ArgE purification on a S200 column. **B)** 12% SDS-PAGE gel of untagged ArgE purification: lane 1 LMWM (M), lane 2 CFE (L), lane 3 insoluble fraction (P), lane 4-7, lane 8 sample after overnight dialysis (D) and lane 9-13 S200 fractions (3 mL each).

An activity assay on non-tagged ArgE was performed *via* the high-throughput L-AAO assay. The kinetic parameters calculated for *N*-Ac-L-Met (Figure 3.38 B) match the values obtained for the N-HisTag ArgE construct. The presence of a His tag at the N-terminus of ArgE, does not influence the enzyme activity and stability, a problem found observed for other metallo-proteins.[174–176]

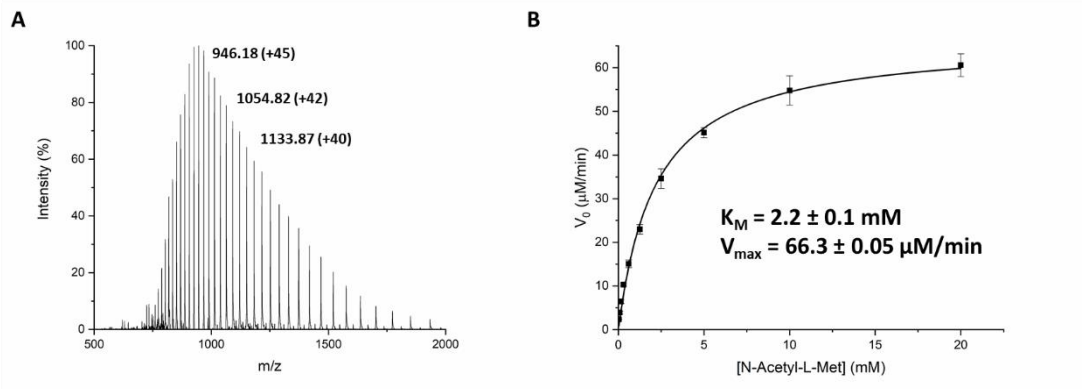


Figure 3.38 Characterization of purified untagged ArgE, following N-His tag cleavage.**A)** Denaturating LC ESI-MS of untagged ArgE (20 µM) with a mass of 42533.08 ± 1.42 Da, in accordance with the MW calculated from the sequence (42532.53 Da). The values reported are m/z with the charges in brackets. **B)** Michaelis-Menten plot of untagged ArgE for *N*-Ac-L-Met. Initial rates calculated from the linear portion of the absorbance curves and fitted with the OriginLab software. The reaction mixture contained: various substrate concentrations, ArgE 10 µg/mL, $CoCl_2$ 100 µM, L-AAO 70 mU/mL, HRP 10 U/mL and *o*-dianisidine 0.1 mg/mL in NaPi 0.1 M pH 8. The plate was incubated at 40°C for 1 h

and absorbance measured at 436 nm every 30 s. Reactions repeated in triplicates, error bars calculated as the standard deviations.

Freshly purified protein at a concentration of 11 mg/mL was employed for crystallization screening. The crystal plates were incubated at 18°C and after three days protein crystals, with a singular prism shape (Figure 3.39), formed for two of the conditions screened: Na formate 0.8 M + TRIS 0.1 M pH 8 + 10% polyethylene glycol (PEG) 8K + 10% PEG 1K and Na formate 0.8 M + TRIS 0.1 M pH 8 + 8% PEG 20K + 8% PEG 550 MME. These crystals were fished, frozen in cryoprotectant and sent to the Diamond Light synchrotron, Harwell for X-ray diffraction analysis. However, the quality (6 Å) and reproducibility of the crystals were very poor. To improve the quality and reproducibility of the protein crystals, non-tagged ArgE E144A was prepared.

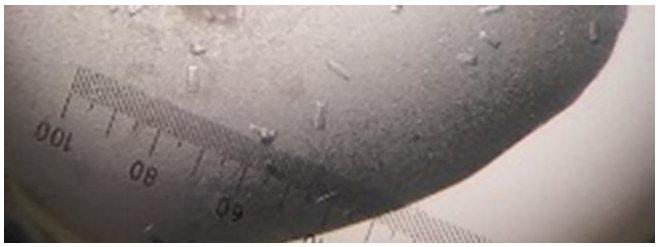


Figure 3.39 Untagged ArgE crystals obtained from a 1 to 1 dilution of freshly purified enzyme at 10 mg/mL in Na formate 0.8 M + TRIS 0.1 M pH 8 + 10% PEG 8K + 10% PEG 1K.

3.11.4 Crystal Optimization

The M20 peptidase family members undergo a significant conformational change upon substrate binding and during catalysis,[169] which may impact the crystals resolution and reproducibility. However, due to the lower activity of the E144A mutant (250-fold less than the WT), coupled with its high affinity for *N*-Ac-AAs, this mutant may yield diffraction quality crystals of higher resolution with the substrate bound.

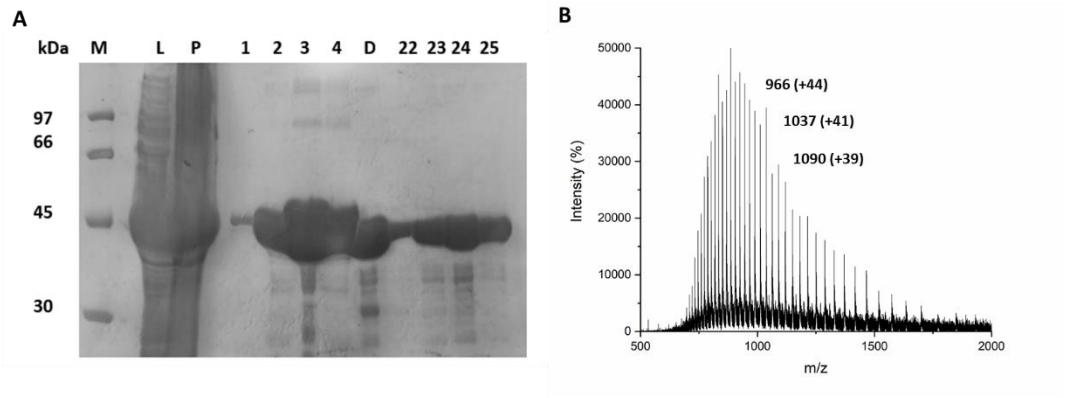


Figure 3.40 Purified untagged ArgE E144A characterization **A)** 12% SDS-PAGE gel of untagged ArgE E144A purification: lane 1 LMWM (M), lane 2 CFE (L), lane 3 insoluble fraction (P), lane 4-7, lane 8 sample after overnight dialysis (D) and lane 9-12 S200 fractions (3 mL each). **B)** Denaturating LC ESI-MS of untagged ArgE (20 µM) with a mass of 42476.02 ± 1.86 Da, in accordance with the MW calculated from the sequence (42474.49 Da). The values reported are m/z with the charges in brackets.

Non-tagged ArgE E144A was purified as described for the non-tagged ArgE WT. SDS-PAGE and denaturing LC-MS analysis show that the correct construct had been expressed (Figure 3.40 A and B). Freshly purified enzyme at a concentration of 15 mg/mL was employed to set up 48 wells crystallization plates *via* the hanging drop method, using the conditions found whilst screening the non-tagged ArgE WT.

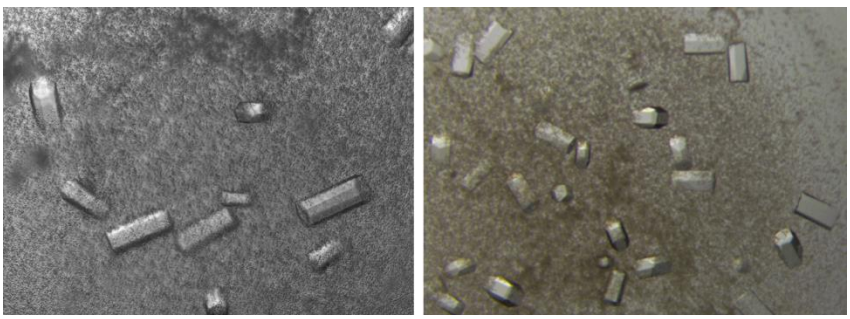


Figure 3.41 Picture of untagged ArgE E144A, crystals obtained from a 1 to 1 dilution of freshly purified enzyme at 15 mg/mL in Na formate 0.8 M + TRIS 0.1 M pH 8 + 10% PEG 8K + 10% PEG 1K with the hanging drop method (drop size 4 µL).

Prism shaped protein crystals (Figure 3.41) were obtained in both plates and sent to the Diamond facility; unfortunately, the best resolution obtained was ~4.0 Å. The DapE crystal structure (PDB entry code 3ISZ and 3IC1) was used as model for molecular replacement with the X-ray data available (Chapter 8.14). However, according to Prof. Grogan the electron density map is too disperse to properly resolve the ArgE crystal structure.

During crystallization a fraction of the enzyme precipitates from the solution, this precipitate forms a glue-like film over the drop (Figure 3.41); this film makes crystal fishing quite challenging and may be responsible for the poor crystal quality. To minimize the precipitation and improve the crystal resolution various: purification buffers, cosolvents, additives, ligands and metal ions were tested, but no improvement was obtained. Without a crystal structure available a random mutagenesis approach was employed to engineer ArgE. In conclusion for the first time we reported the formation of ArgE crystals, however to resolve the enzyme structure the crystal quality needs improvement. A possible solution is to change structure, maybe truncate the enzyme or add solubility tags or try to improve stability *via* mutations and surface entropy reduction.

3.12 Engineering of ArgE *via* random mutagenesis

In the absence of a crystal structure error prone PCR (ep-PCR)[48,177] was employed in an effort to create an active mutant towards *N*-Ac-PT. Random mutagenesis techniques generate large mutant library and a high-throughput assay is required to efficiently screen all variants for activity (Figure 3.42).

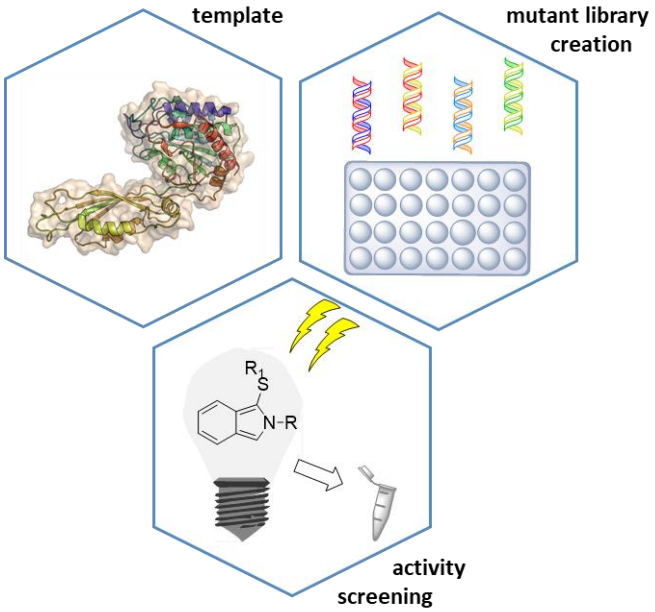


Figure 3.42 Schematic representation of a directed evolution process: from the template protein a mutant library is generated, and the mutants are expressed on a small scale and screened for activity. A high-throughput assay is employed to monitor product formation.

However, none of the assays available can be employed for this task. The development of a new high throughput assay is needed.

3.12.1 *In vivo Arabidopsis thaliana* assay

One option would be to test the ArgE mutants activity using an *in vivo* assay, since L-PT is a potent herbicide, while no plant toxicity has been reported for **1a**. *Arabidopsis thaliana* are often employed for these toxicity assays since they grow in a relatively short period and do not require particular conditions. This work was developed on placement at Syngenta (Jealott's Hill, UK).

A control plate was prepared with *Arabidopsis* seeds incubated with various concentrations of racemic **1a** and **2a**, mixed with minimal agar. The plate was incubated at 20°C with an 8 h light/ 16 h darkness cycle. As shown in Figure 3.43, at increasing D,L-PT concentrations an increase in plant death can be clearly observed, while no toxicity was observed even at the highest *N*-Ac-PT concentration employed.

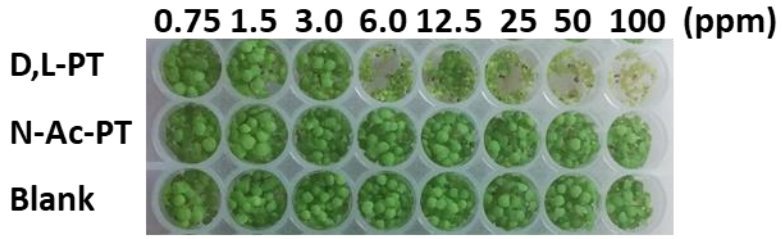


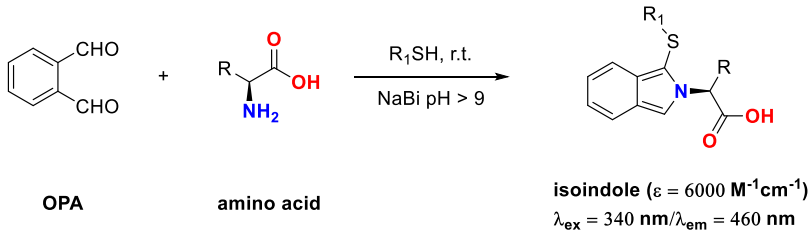
Figure 3.43 Picture of the Arabidopsis assay control plate after one-week incubation at 20°C. Comparison between racemic PT and *N*-Ac-PT induced toxicity at increasing compounds concentration. Blank samples contained NaPi buffer 0.1M pH 8.

The *Arabidopsis* assay is reliable and robust with a good limit of detection (LOD) around 3-6 ppm; however, a week is required for the plant to grow. Given the long incubation time an alternative was sought.

3.12.2 OPA Assay

OPA,[154] along with ninhydrin, is one of the oldest derivatization agents used for amino acid analysis. The OPA reagent readily reacts with any free primary amine groups, however is completely inactive towards *N*-protected amino acids, making this assay ideal for monitoring the hydrolysis of *N*-Ac-AAs. The primary amines group of an amino acid reacts with OPA in the presence of a thiol, 2-mercaptoethanol or *N*-acetyl cysteine for instance, at basic pH; the

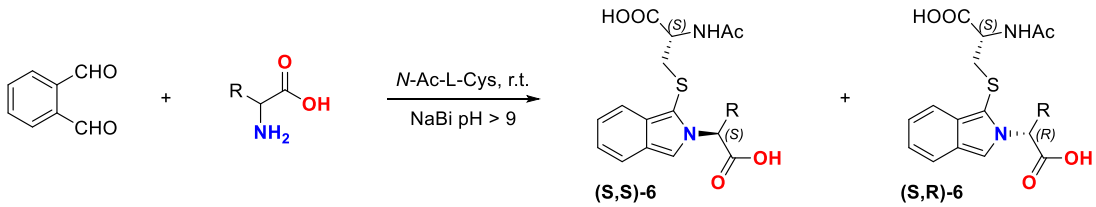
reaction is instantaneous and the isoindole product has a strong absorbance at 340 nm and fluorescence at 460 nm (Scheme 3.15).[178]



Scheme 3.15 OPA derivatization reaction in the presence of an L-amino acid and a thiol at basic pH for the formation of a fluorescent isoindole derivative.

This protocol is advantageous for *pre*- or *post*-column derivatization and HPLC analysis, since the derivatives have the same extinction coefficient (6000 $M^{-1}cm^{-1}$), independently of the amino acid side chain. The main drawback of this assay is the short half-life of the isoindoles, which are dependent on the nature of the functional groups on the amino acid chain, the pH of the solution and the thiol used.[179]

In 2019 a new OPA derivatization protocol, optimised for the high-throughput screening of PT **2a**, was published.[178] The authors used the -SH group of *N*-Ac-L-Cys to catalyse the isoindole formation and since the thiol employed is enantiopure, it will form a diastereomeric mixture upon reacting with a racemic amino acid, as reported in (Scheme 3.16). Similarly, to MR derivatization, this OPA assay protocol allows for a rapid determination of the amino acid enantiopurity, since the diastereoisomers formed can easily be separated on a C18 column by HPLC analysis.



Scheme 3.16 Diastereoselective version of the OPA assay; enantiopure *N*-Ac-L-Cys is used as thiol to produce a 50:50 mixture of two diastereoisomers (R,S and S,S) upon reaction with a racemic amino acid.

The isoindole derivatives of PT were prepared using the optimised protocol reported by Zheng *et. al.*[178] and their stability was tested by monitoring the absorbance at 340 nm of the reaction mixture with a micro-plate reader; no variations in Abs intensity were recorded for over one hour at r.t. (Figure 3.44 A). Unfortunately, the linear range of the assay is quite

narrow (Figure 3.44 B), thus its application for quantitative analysis is limited, but it still is very effective to measure a relative increment in enzymatic activity.

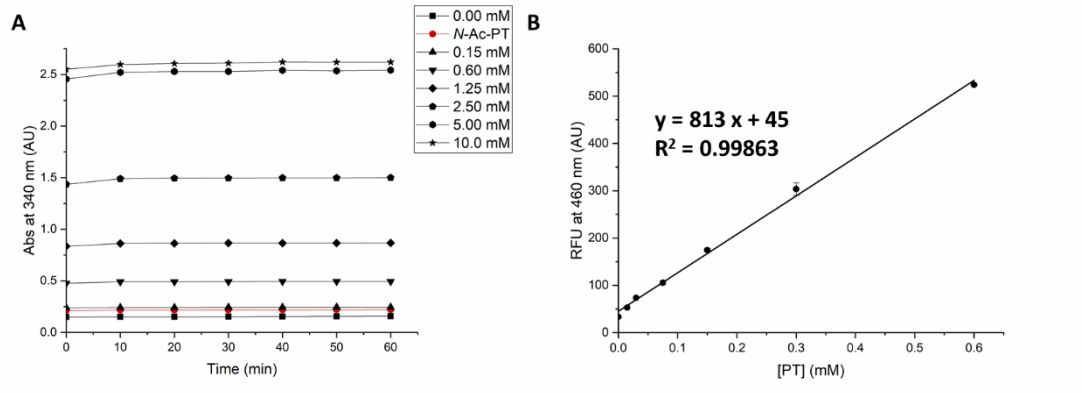


Figure 3.44 Formation of amino acid isoindoles derivatives *via* OPA assay. **A)** Absorbance at 340 nm over time of OPA-PT derivatives. **B)** OPA assay calibration curve for PT, fluorescence intensity recorded at 460 nm (excitation at 340 nm). The OPA assay mixture contained 15 mM OPA and 30 mM *N*-Ac-L-Cys diluted in 50 mL of NaBi 0.14 M pH 9.5. Reactions repeated in triplicates, error bars calculated as the standard deviations.

The optimised OPA assay protocol was employed for the screening of the mutant library.

3.12.3 ArgE mutant library creation and screening

Of the possible mutagenesis methods we selected ep-PCR. ep-PCR was employed to insert point mutations into the *argE* gene at random positions. The amplified gene library was ligated back into the pETHISTEV plasmid backbone *via* MEGAWHOP PCR[180,181] (Figure 3.45) and the reaction mixture was transformed with BL21 (DE3) *E. coli* cells for screening, however this protocol yielded a low number of colonies.

Figure 3.45 Two-step protocol for the creation of ArgE mutant library *via* EP- and MEGAWHOP-PCR. The *argE* gene (blue) is amplified by ep-PCR to generate a library of mutants (red, green, light blue), which are ligated back into the expression vector backbone (yellow) *via* MEGAWHOP-PCR.

Any attempts in optimising the MEGAWHOP PCR step was not fruitful, hence a new ligation protocol was employed Gibson assembly (Figure 3.46).[182]

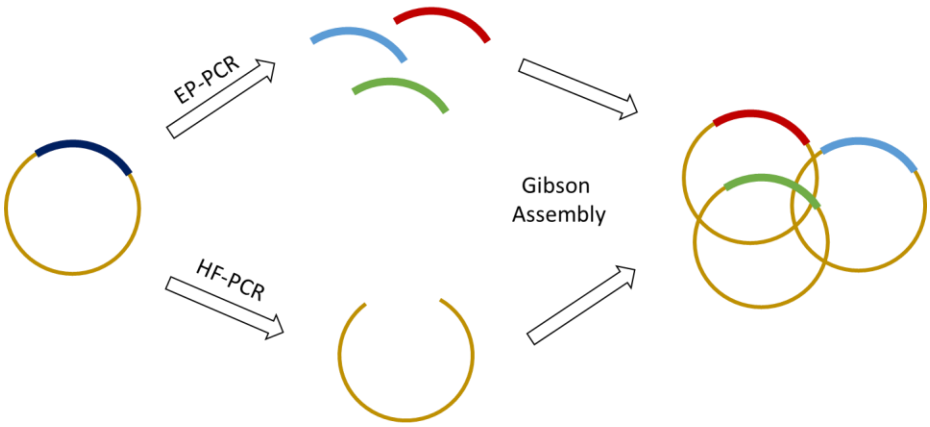


Figure 3.46 Three-step protocol for the creation of ArgE·X mutant library. ep-PCR and HF-PCR employed to insert random mutations in the *arge* gene and amplify the vector backbone, respectively. The gene library is ligated into the vector in the third step *via* Gibson assembly.

Two sets of primers with overlapping ends (10-20 bp) were designed.The first set was used to amplify the *arge* gene *via* EP-PCR to create the mutagenesis library. The second set was employed to amplify the pETHISTEV plasmid backbone with high-fidelity PCR. The two PCR products were purified and mixed with the Gibson assembly master mix to ligate the gene back into the backbone in a tri-enzymatic process: the exonuclease degrades the DNA fragments at the 5' end, the single helix fragments binds to each other thanks the overlapping portions on their sequence and the DNA polymerase start to add the nucleotides; finally, the ligase binds the fragments and seal the nicks in the sequence (Figure 3.47).

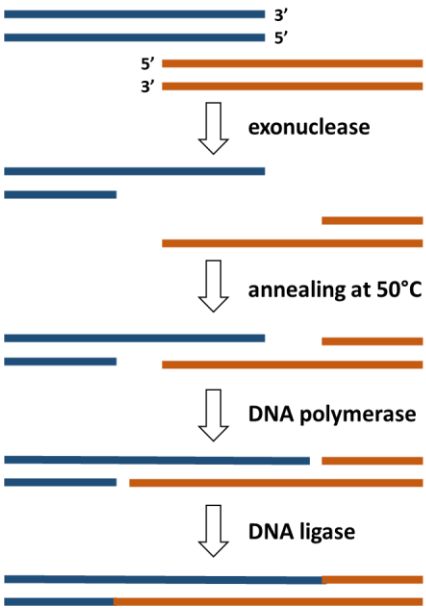
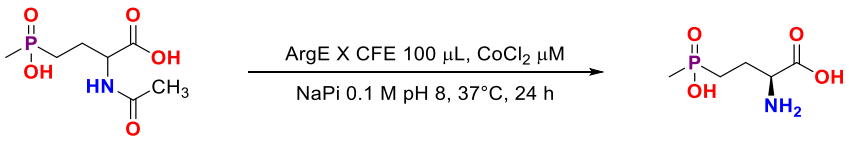


Figure 3.47 Gibson assembly mechanism for the ligation of two DNA strands with overlapping segments.

A sample of the Gibson assembly reaction mixture was transformed and the BL21 (DE3) *E. coli* colonies (~200) used for screening. The colonies were picked and the mutant library (ArgE·X) grown on small scale (1 mL LB media) in 48 well sterile plates at 37°C, until OD_{600} = 0.6-1.0; protein expression started by addition with 0.2 mM IPTG and 0.1 mM $ZnSO_4$. The plates were incubated for 20 h at 20°C and harvested by centrifugation, cell free extract was prepared by addition of BugBuster reagent (primary amine free).

ArgE·X CFE activity was screened against substrate **1a** (2 mM) by setting up analytical scale reactions in a 96 well plate (V_f = 200 µL), as shown in Scheme 3.17.

After 24 h incubation a sample of the reaction mixture was derivatised *via* OPA assay and the isoindole's fluorescence at 460 nm (λ_{ex} = 340 nm) was recorded, unfortunately no increase in activity was detected.



Scheme 3.17 Reaction conditions used during ArgE·X screening foe **1a** activity.

Despite the negative results obtained from the first screening, an efficient protocol for ArgE directed evolution was optimised. However, more time is required to create and screen mutants for *N*-Ac-PT activity.

3.13 Conclusion and Future Work

This work proves that *E. coli* ArgE has great potential for biocatalytic application thanks to its high TONs, stereoselectivity and wide substrate scope for both *N*-acetylated canonical and NCAAs; unfortunately, no activity was detected for *N*-Ac-PT target and protein engineering along with structural studies are required to increase the enzyme affinity for this particular substrate.

The N-His tagged ArgE construct (pETHISTEV-ArgE) was successfully expressed, purified and characterised; ICP-MS and EDXRF[146] confirm the presence of only one Zn^{2+} ion bound for ArgE monomer, while any attempt to directly express the enzyme bound to Co^{2+} was unfruitful.

A series of assays were tested to find the best protocol to observe the amino acid formation in solution and calculate ArgE catalytic activity. The MR protocol was quite effective for the analysis of standards solution, unfortunately the results were not reproducible when employed for the derivatization of enzymatic reaction mixtures. Hence the chiral HPLC assay[121] was exploited to study the stereoselectivity of the enzyme by either observing the direct formation of the product (amino acid with an UV-Vis absorbance) or the disappearance of the starting material (amide bond absorbance at 210 nm) in a kinetic resolution. However, to quantify the reaction conversion, an efficient ^{1}H-NMR assay was developed; by recording a ^{1}H-NMR spectrum of the reaction mixture, the relative conversion of the biotransformation can be easily obtained by comparing the integrals of the CH_3-group for the free and *N*-bound

acetate. This assay can be employed to directly monitoring hydrolysis and to optimise reaction conditions, however, since the ArgE rate limiting step is the H_2O hydrolysis and the assay requires the use of D_2O as solvent, the enzyme kinetic parameters cannot be accurately determined *via* ^{1}H-NMR spectroscopy. Instead the L-AAO assay reported by Dr Sànchez-Carròn[122] was employed in the kinetic and activity study of ArgE for both *N*-acetyl proteinogenic and non-proteinogenic amino acids, thanks to the wide substrate scope of the oxidase.

Activity and stability (TDA) studies showed that the best reaction conditions for the enzyme are: neutral or basic pH, a temperature of 40°C and addition of $CoCl_2$. Even if not classified as thermostable (T_m of 60°C, at neutral and basic pH), the *E. coli* ArgE still retains 96% of its activity at 50°C, showing potential to be coupled with the thermostable NAAAR in a KDR. While Co(II) ions cause a 1.4-fold increase in enzyme activity, the addition of $ZnCl_2$ strongly inhibits the enzyme, with Zn^{2+} acting as a reversible competitive inhibitor; a similar case was already reported for at least one other zinc dependant hydrolase.[147]

Racemic PT was chemically acetylated to form its *N*-acetylated derivative **1a** in good yields (73%);[83] this protocol was also employed for the acetylation of non-canonical amino acids to create a library of substrates, which was subsequently screened for activity with ArgE. The results obtained with the ^{1}H-NMR and chiral HPLC assays show that the enzyme possesses a wide substrate scope and high stereoselectivity towards the L-isomer. The best conversion percentages were obtained by increasing the loading of ArgE up to 1 mg/mL, however even at high concentration only a 10% conversion was calculated for **1j** and no activity was detected for **1a**.

Activity studies conducted with various *N*-Ac-PT structural analogues suggest that the enzyme substrate binding site is able to bind polar groups, such as the sulfoxide and sulfone in **1b** and **1c**, but does tolerate the presence of a negative charge on the amino acid side chain (the phosphinic and carboxylic acid of **1a** and Glu).

With a lack of structure, we are not sure which residues are involved in the recognition of the substrate side chain. Protein engineering is required to obtain an ArgE variant active towards *N*-Ac-PT. Various ArgE constructs were expressed and purified (N-HisTag ArgE WT, C-HisTag ArgE WT, non-tagged ArgE WT and non-tagged ArgE E144A) to screen for crystal formation; prism shaped protein crystals were obtained using the two non-tagged constructs with the best resolution (4 Å) achieved using the mutant enzyme. The crystal structure of

ArgE's closest homologue (H. influenzae DapE, PDB 3IC1) was employed to resolve ArgE structure *via* molecular replacement, unfortunately the electron density map was too dispersed. The ArgE crystal quality needs to be improved to obtain a better dataset; a surface entropy reduction strategy may prove fruitful in this instance.

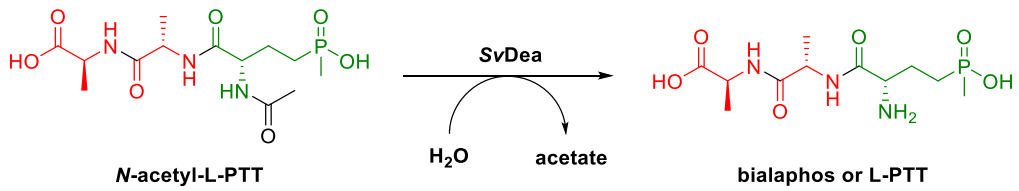
Sequence alignment and homology model studies were employed to identify various residues involved in catalysis, cofactor binding and substrate stabilization in ArgE. These residues were mutated into alanine *via* SDM, expressed, purified and their activity tested with the L-AAO assay. E144, H195 and R264 appear to play a fundamental role in the reaction mechanism, as previously suggested by Holz[169] with the glutamate acting as general acid/base catalyst, the histidine activating the acetate carboxylate *via* hydrogen bond formation and the arginine stabilizing the substrate binding by interacting with the amino acid α-carboxylate.

Since the ArgE crystal structure was not available a random mutagenesis strategy was employed to evolve the enzyme. An optimised directed evolution protocol (EP-PCR and Gibson assembly)[177,182] along with the high-throughput OPA assay[178] were employed to create and screen a ArgE·X mutant library for **1a** activity. Unfortunately, no improved activity variants were isolated and more time is required to efficiently generate and screen a large number of mutants.

4 The *N*-acetyl-bialaphos deacetylase (*Sv*Dea)

Unfortunately, the *Escherichia coli* ArgE was inactive towards *N*-Ac-L-PT and attempts to engineer the enzyme by rational design or directed evolution were not fruitful. However, this deacetylase proved to have good potential for the KR of *N*-acetyl non-canonical amino acids (NCAAs).

Therefore, a second hydrolytic enzyme was selected as a candidate for the DKR of N-Ac-L-PT; the *Streptomyces viridochromogenes* *N*-acetyl bialaphos deactylase (*Sv*Dea) has been proposed to catalyse the last step in the synthesis of the natural herbicide tripeptide bialaphos (L-PTT), the deacetylation of the *N*-acetyl-L-PTT precursor (Scheme 4.1). The gene encoding *Sv*Dea is found within the operon, which contains 16 genes involved in the bialaphos biosynthesys (Figure 1.5 and Table 1.1). This enzyme has been described in the literature,[23,25,26,183] however it has never been isolated or characterised, thus no background information on substrate specificity was available. Since most hydrolases display quite a wide substrate scope and can accept non-natural substrates, *Sv*Dea may be able to efficiently catalyse the hydrolysis of *N*-Ac-L-PT.



Scheme 4.1 The predicted last step in the biosynthesis of the non-ribosomal tripeptide bialaphos, the *Sv*Dea catalysed deacetylation of *N*-Ac-L-PTT.

4.1 Aims

The aims of this part of the overall project were to express, purify and characterize the novel *Streptomyces viridochromogenes* Tü494 *Sv*Dea and to assay its activity for the deacetylation of the natural herbicide L-PTT and for the NAAAR DKR of L-PT. To achieve this goal the optimised ^{1}H NMR assay described in the previous chapter was employed to monitor acetate production, along with LC-ESI MS analysis; while a colorimetric assay was used to determine the kinetic parameters of the enzyme. Sequence alignment, homology modelling and crystal

trials were also performed to obtain some structural and mechanistic insights of this novel deacetylase.

4.2 Expression and purification of N-His tagged *Sv*Dea.

The full length *S.v*. *dea* gene, cloned into the pUC57 plasmid was ordered from GenScript (Uniprot code: Q56171). The gene was subcloned into the expression vector pET28a using the NdeI and XhoI restriction sites, the N-hexahistidine tagged pET28a-*Sv*Dea construct was transformed in *E. coli* BL21 (DE3) cells and a series of expression tests were carried out to select the best media and optimise the IPTG concentration used for the enzyme production. All samples were analysed *via* SDS-PAGE. A low yield of protein expression was obtained using three different media: LB, sorbitol media and sorbitol media prepared in 0.1 M phosphate buffer pH 7.8;[184] however, overnight expression (~20 h) in terrific broth (TB) with 0.4 mM IPTG yielded a good amount of soluble *Sv*Dea (Figure 4.1). These optimised conditions were employed for scale-up expression.

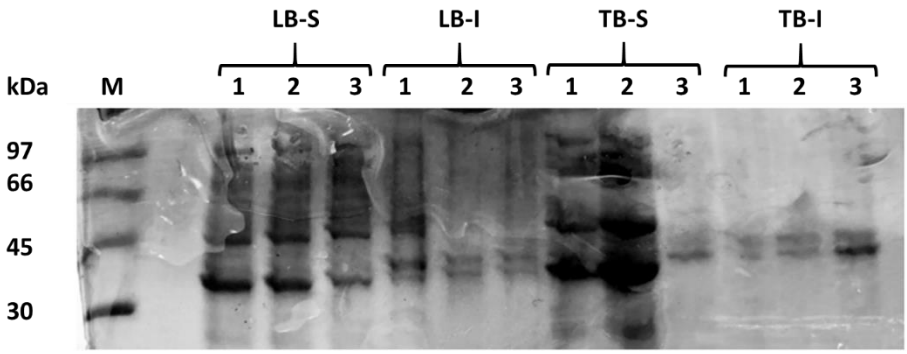


Figure 4.1 15% SDS-PAGE analysis of N-His tagged *Sv*Dea expression tests. Lane 1 low molecular weight marker LMWM (M), lane 2-7 LB-media soluble (S) and insoluble (I) samples and 8-13 lane TB-media soluble (S) and insoluble (I) samples. Protein expression was induced with 0.1 mM, 0.2 mM and 0.4 mM IPTG for samples 1, 2 and 3, respectively with incubation at 20°C for 20 h. The *Sv*Dea band is at ~ 35 kDa.

For *Sv*Dea purification we used the two-step protocol described for the isolation of N-His tag ArgE; the harvested cells were resuspended in lysis buffer (Tris 50 mM pH 8) and lysed *via* sonication treatment for 15 minutes (30s ON/30s OFF. These steps were particularly tricky, given the gelatinous consistency of the cells it was not possible to obtain a homogenous suspension; furthermore, following lysis a clouded solution was obtained, instead of a clear dark lysate. This suspension was centrifuged, filtered and loaded on a HisTrap Fast Flow 1 mL column. N-His tagged *Sv*Dea eluted at 140 mM imidazole (Figure 4.2 A); analysis *via* SDS-PAGE electrophoresis and denaturing LC-ESI MS of the collected fractions confirmed the isolation

of the correct enzyme with an observed MW of 34201.25 ± 0.33 Da (Figure 4.2 B), which was in accordance with the theorical MW minus the initial methionine: 34203.39 Da, calculated from the protein sequence with the ExPaSy ProtParam tool. Unfortunately, the final yield of expressed enzyme after just a single purification step was very low: 1 mg *per* L of culture.

However, the expression test SDS-PAGE (Figure 4.1) indicates the production of a good level of soluble protein; thus, the loss of enzyme must occur during one or more purification steps. Before testing the enzyme's activity, further optimization of the CFE preparation and purification protocols are required to increase the yield of soluble protein isolated.

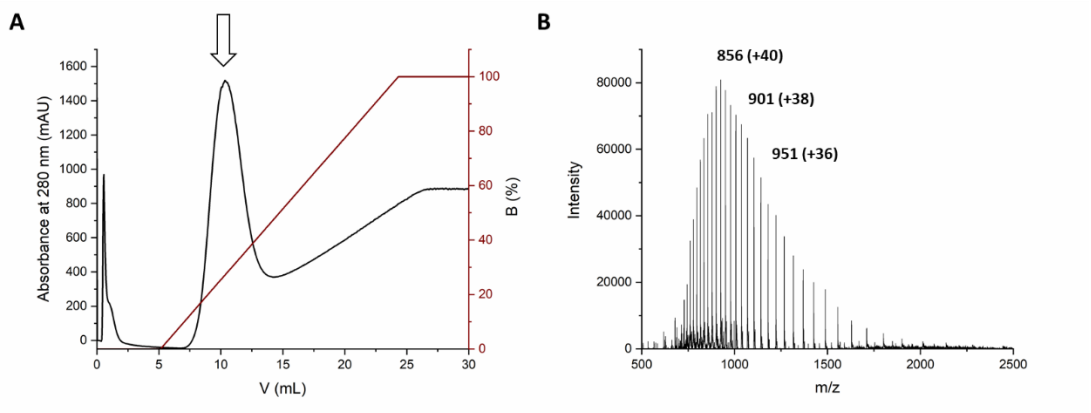


Figure 4.2 Purification and characterization of *Sv*Dea. **A)** IMAC chromatogram for N-His tagged *Sv*Dea on an FF HisTrap (1 mL) column recorded at 280 nm. The arrow points to the deacetylase peak. **B)** Denaturating LC ESI-MS of *Sv*Dea (20 µM) with a deconvoluted mass of 34201.25 ± 0.33 Da, in accordance with the MW calculated from the sequence (34203.39 Da). The values reported are *m/z* with the charges in brackets.

4.3 Thermal denaturation assay (TDA)

This is the first report of recombinant *S. viridochromogenes* Dea isolation and expression, hence no prior knowledge was available regarding the optimal conditions for the enzyme stability. Since Tris pH 8 is a common buffer used for the purification of various enzymes, including *E. coli* ArgE, it was also selected to purify *Sv*Dea. However, given the low yield of pure protein obtained, a TDA was run to determine the best buffer to employ to increase protein stability during purification. A 0.1 mg/mL *Sv*Dea solution (3 µM) was used to run the assay on a 96 well plate; the enzyme melting curves were recorded in twelve different conditions: 0.1 M Na citrate (pH 3.0, 3.5, 4.0, 4.5, 5.0 and 5.5), 0.1 M NaPi (pH 6.0, 6.5, 7.0, 7.5 and 8.0) and 0.1 M Na carbonate (NaCi) pH 9.0.

As reported in Figure 4.3 A, the melting temperatures of SvDea are consistently low, between 35° and 40°C, across a wide range of pHs, but start to increase in alkali solutions, reaching a T_m of 61°C in 0.1 M NaCi pH 9.0. Considering this first set of results, it was clear that the enzyme stability is higher at basic pHs. A second TDA was run to compare *Sv*Dea's melting temperatures for two alkali buffers, 0.1 M NaCi and 0.1M *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS). A constant T_m of 65°C between pH 8 and 11 was obtained using CAPS, thus this buffer was selected to carry out the purification and store the protein.

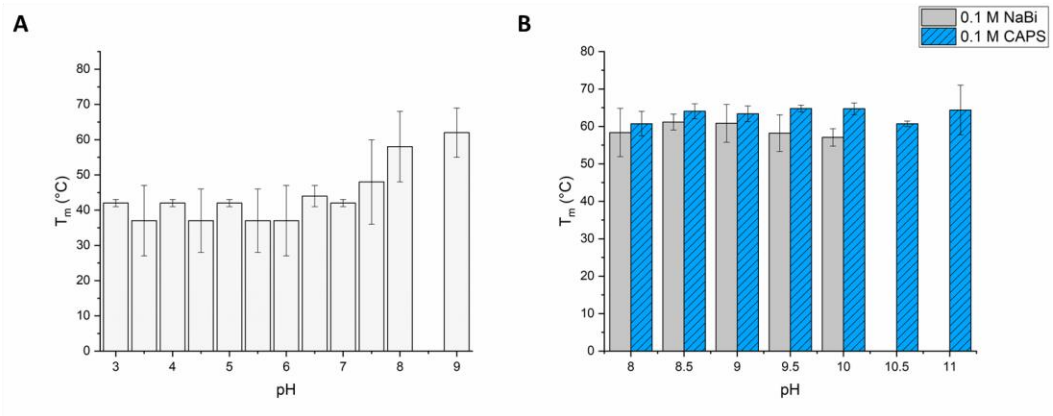


Figure 4.3 TDA analysis of *Sv*Dea. **A)** *Sv*Dea melting temperatures at various pH; the assay mixture contained *Sv*Dea 3.0 µM (0.1 mg/mL), 5X SYPRO orange in DMSO and 50 mM buffer at various pHs. **B)** Comparison between *Sv*Dea melting temperatures at alkali pH using an organic (blue bars) and inorganic (grey bars) buffer; the assay mixture contained SvDea 3.0 µM (0.1 mg/mL), 5X SYPRO orange in DMSO and 0.1 M buffer at various pHs. The melting curves were obtained measuring the fluorescence of SYPRO orange between 25 and 75°C with a gradient of 1°C/min. Reactions repeated in triplicates, error bars calculated as the standard deviations.

4.4 Optimization of the purification of N-HisTag *Sv*Dea

The *Sv*Dea was expressed using the optimised protocol (Chapter 6.3.5), the harvested cells were resuspended in 0.1 M CAPS buffer pH 9.5. Given their gelatinous consistency and the low protein yield obtained previously during purification using sonication, two alternative techniques were tested to break down the cell walls: French press and overnight treatment with lysozyme at 4°C. Precipitation was observed using the high pressure treatment with the formation of a cloudy lysate, while the chemo-lysis method resulted more effective; however, the resulting suspension had quite a dense consistency and could not be efficiently filtered. The overexpression of *Sv*Dea in the cytosol may have tampered with the biosynthesis of the cell membrane or wall components, which would explain the higher viscosity of the CFE. To proceed with the purification *via* IMAC the lysate was incubated for 2 h with Ni^{2+} resin

at 4°C on an orbital shaker (150 rpm) and loaded onto a column; the protein was eluted manually with 0.1 M CAPS buffer pH 9.5 with 500 mM imidazole. Bradford assay was used to determine which fractions contained the enzyme (Figure 4.4 A), these were collected, concentrated and loaded on a Superdex S75 column.

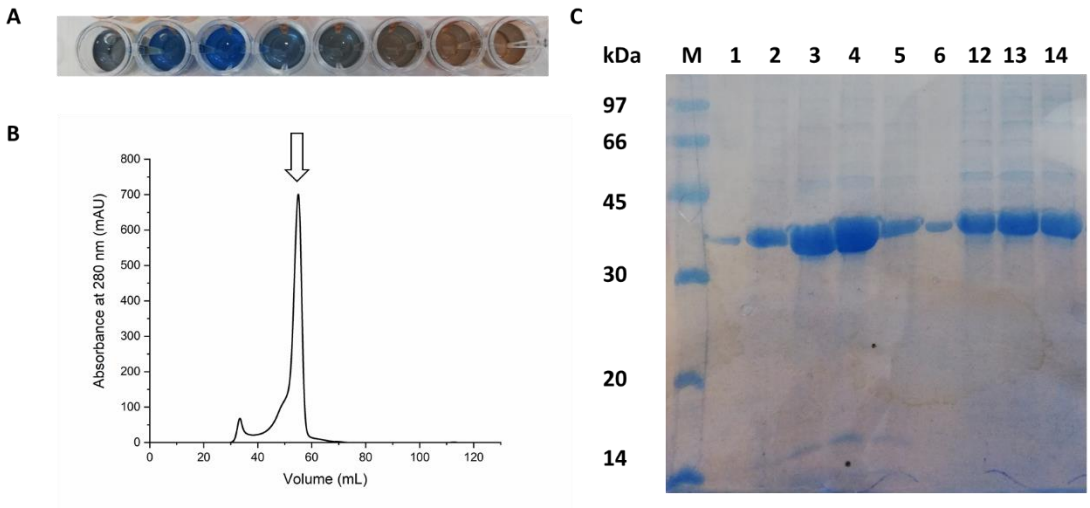


Figure 4.4 Isolation of *Sv*Dea **A)** 96-well plate used to run a qualitative Bradford assay to identify the IMAC fractions containing *Sv*Dea, tubes 1-6 (blue) were collected and used to carry out the second purification step. **B)** Size exclusion chromatogram for N-His tag *Sv*Dea on a Superdex S75 column. The arrow points to the deacetylase peak. **C)** 15% SDS-PAGE gel of N-HisTag *Sv*Dea purification: lane 1 LMWM (M), lane 2-7 manual Ni^{2+} resin column fractions (~7 mL each) and lane 8-10 S75 fractions (3 mL each).

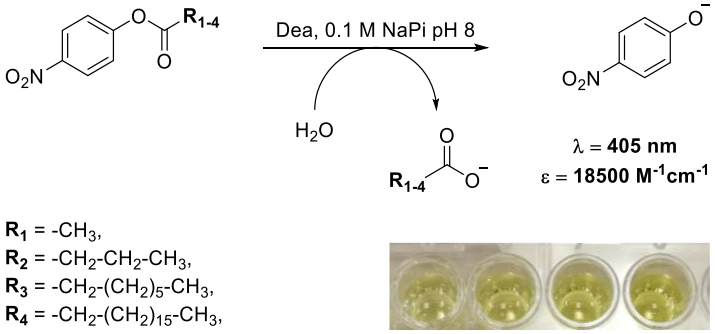
The SEC absorbance chromatogram, monitoring at 280 nm, shows a single peak at an elution volume of ~55 mL (Figure 4.4 B), which corresponds to a molecular weight of 35 kDa, this suggest the presence of a monomeric hexa-His tagged *Sv*Dea in solution. The enzyme was obtained with a purity of >95% (Figure 4.4 C) and an average yield of 15 mg *per* L of culture. Both SDS-PAGE analysis and denaturing LC-ESI MS analysis confirm that the desired construct was expressed and isolated.

4.5 Assaying *Sv*Dea activity and kinetic analysis

4.5.1 The *p*-NO_2 phenol esterase assay

A sequence homology study revealed that *Sv*Dea shares between 25-50% identity with four esterases, belonging to the hormone-sensitive lipase family (EC 3.1.1.79, HSL). A common colorimetric assay used to measure these enzymes activity is the *p*-NO_2 phenol (*p*NP) assay[185] (Scheme 4.2), where various *p*-NO_2 phenol esters are employed as substrate for the enzyme.

Following hydrolysis, *p*-nitro phenol is released in solution, which is deprotonated in alkali conditions (pH > 7.8); this negatively charged form is characterised by a bright yellow colour (λ_{max} = 405 nm, ε = 18500 M^{-1} cm^{-1}), easily detected *via* UV-Vis spectroscopy.



Scheme 4.2 Reaction scheme of the *p*NP continuos assay for *Sv*Dea hydrolysis of four *p*NP esters. The deprotonated form of *p*-NO_2 phenol has a bright yellow colour and a strong absorbance at 405 nm, which can be easily monitored *via* UV-Vis on a 96-wells microplate reader.

The continuous *p*NP assay was employed to test *Sv*Dea activity towards four *p*-NO_2 phenol esters: acetate (*p*NPA), butyrate (*p*NPB), octanoate (*p*NPO) and palmitate (*p*NPP). Following 20 minutes incubation at 25°C, only the *p*NPA reaction mixture developed a yellow colour (Figure 4.5 A). This suggests that the enzyme is very specific towards the acyl moiety of the substrate, with only the shorter acetate ester being recognised by *Sv*Dea. Furthermore, the *p*NP assay was employed to determine the best temperature for *Sv*Dea activity, which was monitored at 25°, 30°, 40° and 50°C; four sets of reactions containing *p*NPA, each incubated at a different temperature, were prepared and their absorbance measured over time at 405 nm, the specific activities were calculated dividing the maximum rate (V_{max}) of the hydrolysis with the enzyme loading employed (mg). According to Figure 4.5 B, *Sv*Dea's activity is not particularly affected by the reaction temperature, however the highest value (2100 ± 100 U/mg) was reached at 50°C.

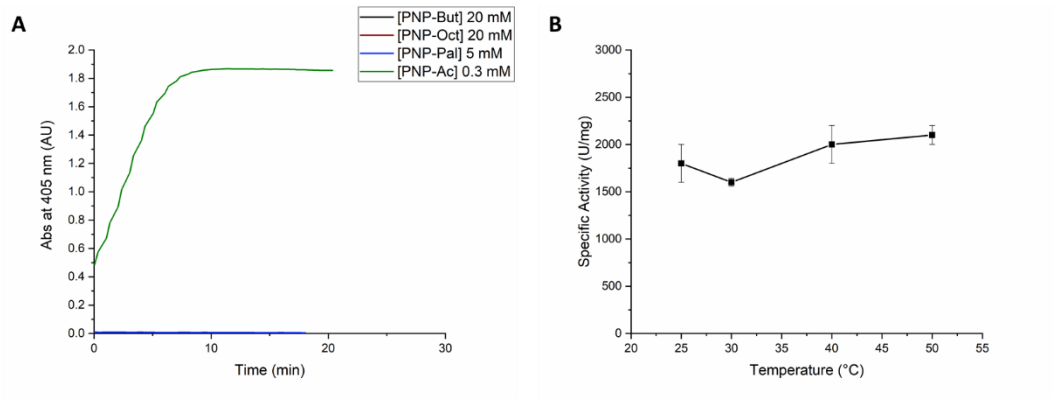


Figure 4.5 *Sv*Dea activity test *via* *p*NP assay. **A)** Comparison between the absorbance curves at 405 nm for the enzymatic hydrolysis of *p*NPA (green), *p*NPP (blue), *p*NPO (red) and *p*NPB (black). **B)** Specific activity of *Sv*Dea for *p*NPA at 25°, 30°, 40° and 50°C measured *via* the continuous *p*NP assay. The reactions contained various substrate concentrations and 0.1 mg/mL *Sv*Dea in NaPi 0.1 M pH 8, plate incubated at the desired temperature for 30 minutes and absorbance measured at 405 nm every 30 s. The specific activity is expressed in units *per* milligram of enzyme, with one unit is referring to the µM of substrate converted *per* minute. Reactions repeated in triplicates, error bars calculated as the standard deviations.

The continuous *p*NP assay reported was further employed for the kinetic analysis of *Sv*Dea towards *p*NPA. A set of reactions at increasing substrate concentrations were incubated at 50°C in 0.1 M NaPi buffer pH 8 and the absorbance of *p*-NO_2 phenol was monitored at 405 nm every 30 s for 1 h. Since *p*NPA is prone to auto-hydrolysis in these reaction conditions, the substrate stock was prepared in acetonitrile to increase its stability and various dilutions into the reaction buffer were tested; surprisingly *Sv*Dea resulted quite tolerant to the presence of organic solvent, no decrease of activity was detected up to 10% ACN in the reaction mixture. For the kinetic analysis a set of control reactions (without enzyme) was also prepared and their absorbance measured over time; the final absorbance for each time-point was calculated subtracting the Abs of the enzymatic reaction from the Abs of the control. The initial reaction rates were obtained from the linear portion of the absorbance curves and fitted using the Michaelis-Menten non-linear fit with the Lab Origins software (Figure 4.6 A and B). The calculated kinetic parameters show that *Sv*Dea has a strong affinity (K_M = 250 ± 30 µM) and high catalytic efficiency (k_{cat}/ K_M = 2052.837 ± 0.008 $M^{-1}s^{-1}$) for *p*NPA, with a TON of 0.51 ± 0.04 s^{-1}.

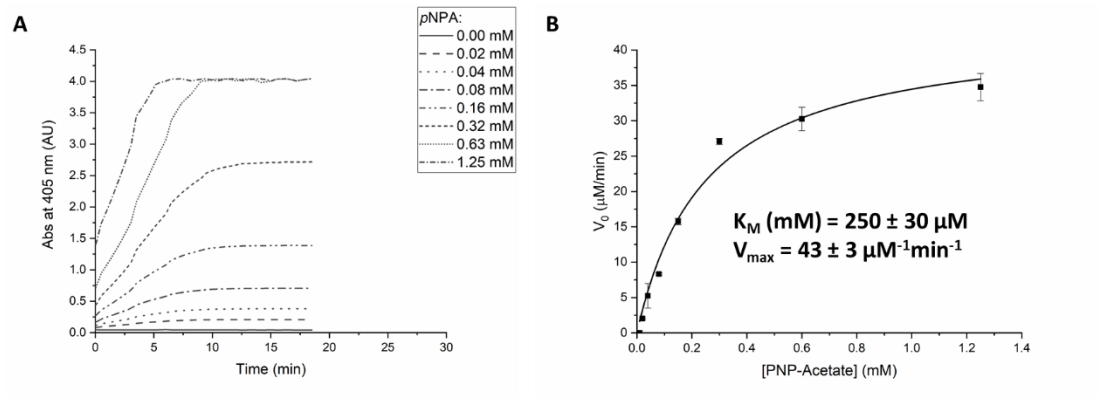


Figure 4.6 *Sv*Dea kinetic analysis *via* *p*NPA. **A)** Typical raw data profile for the formation of *p*-NO_2 phenol over time at various substrate (*p*NPA) concentrations. **B)** Michaelis-Menten plot of *Sv*Dea for *p*NPA. Initial rates calculated from the linear portion of the absorbance curves and fitted with the OriginLab software. The reaction mixture contained: various substrate concentrations and *Sv*Dea 50 µg/mL in NaPi 0.1 M pH 8. The plate was incubated at 50°C for 30 minutes and the absorbance measured at 405 nm every 30 s. Reactions repeated in triplicates, error bars calculated as the standard deviations.

4.5.2 The L-AAO/HRP coupled assay

As well as the *p*NP assay to monitor ester hydrolysis, we also found the L-AAO assay[122] proved to be a quick and efficient method to measure the activity of deacetylases towards a wide range of *N*-acetylated canonical and NCAAs, as reported in Chapter 3. This high-throughput assay was employed to test the hydrolytic activity of *Sv*Dea using ten *N*-acetylated-L-amino acids. Each reaction contained 10 mM substrate and 0.4 mg/mL enzyme; following 1 h incubation at 50°C, no sign of activity was detected in any sample as shown in Figure 4.7, with the exception of the assay positive control (10 mM L-Met). These results suggest that the *Sv*Dea enzyme has a very narrow substrate scope, being extremely selective towards both the amino acid moiety and the acyl chain of the substrate, being able to recognise only the acetylated *p*NP ester.

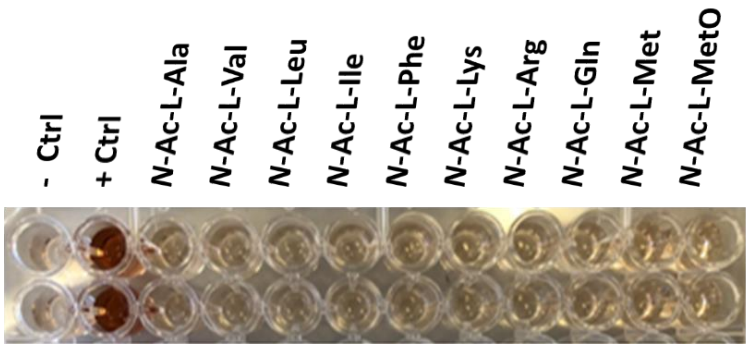


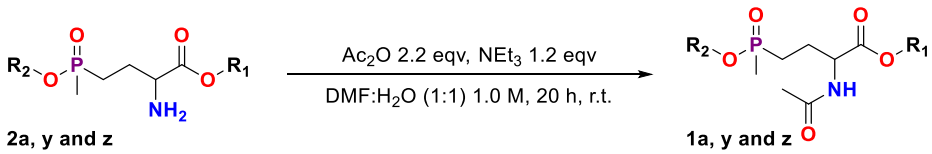
Figure 4.7 Screening of *Sv*Dea activity towards ten *N*-acetylated L-amino acids *via* L-AAO assay. Water and L-methionine used as negative (**- Ctrl**) and positive (**+ Ctrl**) controls respectively. The assay

mixture contained: *N*-acetyl-L-AA 10 mM, *Sv*Dea 0.4 mg/mL, L-AAO 100 mU/mL, HRP 10 U/mL and *o*-dianisidine 0.1 mg/mL in NaPi 0.1 M pH 8. The plate was incubated at 50°C for 1 h and absorbance measured at 436 nm every 30 s.

4.6 Kinetic resolution of *N*-acetyl PT and derivatives

4.6.1 Substrate synthesis

The racemic ethyl 2-amino-4-(ethoxy(methyl)phosphoryl) butanoate (**2y**) and a sample (~2 mg) of the natural product L-PTT (**2z**) were gifted by Syngenta (Jealott's Hill, UK) and acetylated following the same protocol employed for the preparation of **1a** (Scheme 4.3).[83] Ethyl 2-acetamido-4-(ethoxy(methyl)phosphoryl) butanoate (**1y**) was isolated with a yield of 11.5% and characterised by NMR spectroscopy and LC-ESI MS spectrometry, the low yield is caused by the partial hydrolysis of the ester bonds, which spontaneously occur during acetylation. Given the low amount of natural product available a full characterization of *N*-Ac-L-PTT (**1z**) could not be obtained. Following the acetylation work-up, the product was freeze-dried and full conversion of the tripeptide was confirmed by LC-ESI MS analysis; the mass spectrum recorded on a positive ion mode showed two peaks at m/z of 366 and 388, corresponding to the protonated $[M+H]^+$ and sodium adduct $[M+Na]^+$ species respectively.

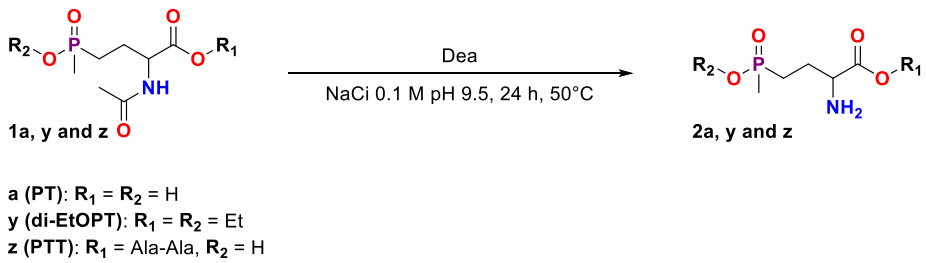


a (PT): $\mathbf{R_1}$ = $\mathbf{R_2}$ = H
y (di-EtOPT): $\mathbf{R_1}$ = $\mathbf{R_2}$ = Et
z (PTT): $\mathbf{R_1}$ = Ala-Ala, $\mathbf{R_2}$ = H

Scheme 4.3 Chemical acetylation protocol for the synthesis of *N*-acetyl-DL-phosphinothricin (**1a**) and derivatives (**1y and z**).

4.6.2 Biotransformations

At first *Sv*Dea activity was tested towards *N*-Ac-PT; a set of four reactions at increasing enzyme concentrations (0.00, 0.25, 0.50 and 1.00 mg/mL) were incubated at 50°C for 24 h (Scheme 4.4). The biotransformations were quenched by addition of 20 µL of $HCl_{conc.}$, centrifuged and the supernatant analysed *via* ^{1}H-NMR spectroscopy and LC-TOF MS; no traces of product formation was detected (Figure 4.8).



Scheme 4.4 *Sv*Dea catalysed hydrolysis of *N*-Ac-PT and derivatives **1a**, **1y** and **1z**. Reactions repeated in triplicates.

The protocol was repeated using compounds **1x** and **1y** as substrates. For the KR of **1x** after 24 h incubation, a mixture of starting material (30%) and partially hydrolysed substrate (70%) was obtained, but no formation of the deacetylated product was detected; however, it is not clear if the ester hydrolysis occurs spontaneously or is mediated by *Sv*Dea. While the formation of traces amount of **2y** were observed by LC-MS analysis of the samples of the reactions of *N*-Ac-L-PTT with 0.50 and 1.00 mg/mL *Sv*Dea. The product peak in the mass spectrum had a low intensity; to obtain more conclusive results, it is necessary to synthesize the natural substrate **1y** in higher quantities and scale up the biotransformations to be able to run a proper NMR characterization.

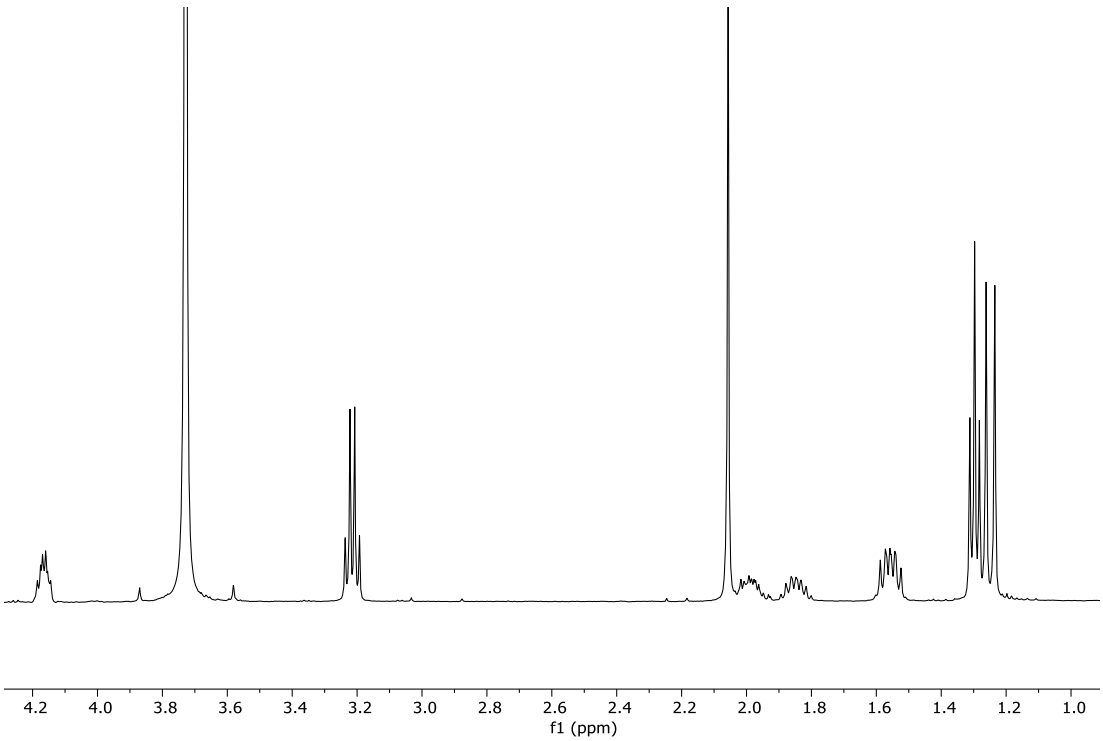


Figure 4.8 ^{1}H-NMR of the *Sv*Dea (1 mg/mL) catalysed deacetylation of 20 mM N-Ac-PT (**1a**). Following 24 h incubation at 50°C (250 rpm), only the substrate was present in the reaction mixture.

Unfortunately, *Sv*Dea cannot be employed for the DKR of L-PT; it appears, based on its proposed role within the biosynthetic pathway, that *Sv*Dea is highly selective towards the intermediate *N*-Ac-L-PTT intermediate, with no activity detected towards any other *N*-acetylated amino acids tested. However, since the enzyme is active towards *p*NPA, it will be interesting to test the enzyme in the deacetylation of other small aromatic esters or derivatives to prove if *Sv*Dea is able to catalyse amide hydrolysis (Figure 4.9).

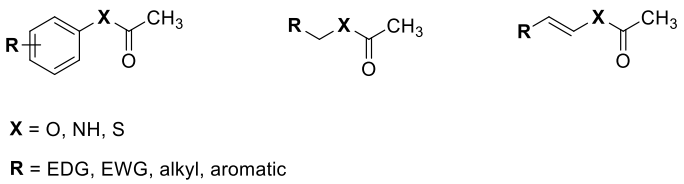


Figure 4.9 Various acetyl ester, thioester or amides with an aromatic, aliphatic or unsaturated chain to test for activity towards *Sv*Dea.

4.7 Structural studies.

4.7.1 Crystal Trials

Given the limited substrate scope of the WT enzyme, the resolution of *Sv*Dea crystal structure could give new insights on the residues involved in ligands recognition and possibly help in designing mutants with higher biocatalytic potential on L-PTT derivatives.

The current pET28a-*Sv*Dea construct available encodes for an N-His tagged enzyme with the *S.v. dea* gene cloned using the NdeI and XhoI restriction sites; to obtain a non-tagged construct the gene was amplified *via* PCR, using the forward plasmid to change the NdeI site into a NcoI. The PCR product was purified *via* agarose gel electrophoresis, digested and ligated into the pETHISTEV vector, which possess a N-terminal TEV cleavable hexa-histidine tag, using the NcoI and XhoI restriction sites.

The pET28a-*Sv*Dea and pETHISTEV-*Sv*Dea constructs were expressed using the optimised *Sv*Dea protocol (Chapter 6.3.5). The N-His tagged *Sv*Dea was purified using the two-step protocol described in Section 4.4. To obtain the untagged enzyme, following IMAC the pETHISTEV-*Sv*Dea fractions were collected and incubated with 2 mg/mL of TEV protease for

16 h at 4°C and dialysed against *Sv*Dea storage buffer (CAPS 0.1 M pH 9.5 and 150 mM NaCl). Unfortunately, the protein completely precipitated during overnight incubation, leaving only the protease in solution, as shown in Figure 4.10.

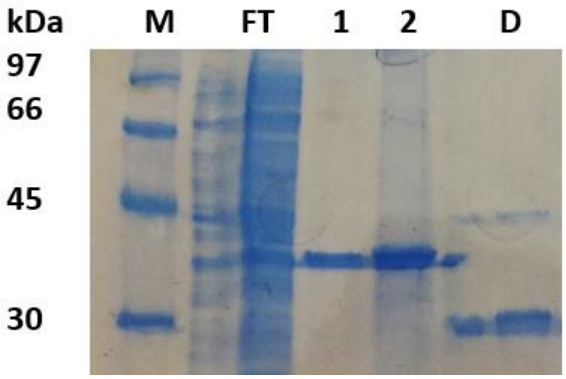


Figure 4.10 15% SDS-PAGE gel of untagged *Sv*Dea purification: lane 1 LMWM (M), lane 2-3 IMAC flow through (FT), lane 3-4 manual Ni^{2+} resin column fractions (~7 mL each) and lane 5 protein sample following overnight incubation with the TEV protease(D).

The initial crystal trials were performed employing the freshly purified N-His tag *Sv*Dea at a 10 mg/mL concentration; ~300 conditions were screened using the crystallization kits: PACT, CSS1+2 and JCSG. Unfortunately, only plate shaped salt crystals were obtained from this initial screening, further studies are required. Trying different crystallization screens and changing the enzyme concentration may deliver more positive results. However, the optimization of the untagged *Sv*Dea purification protocol and possibly the creation of a C-His tag construct may be required to obtain positive hits during crystal trials.

4.7.2 Homology Model

Given the high time-demand required to obtain and optimise protein crystals and the lack of information on this novel deacetylase, a sequence alignment and homology model were built to understand which residues are involved during hydrolysis.

The *Sv*Dea sequence was aligned using the Clustel Omega[186] and ESPrit3[187] softwares with four homologues microbial esterase from the HSL family with a resolved structure: Est22 (27.4 % sequence identity, PDB entry code 5HC0), *Pyrobaculum calidifontis* PestE[188] (33.8 % sequence identity, PDB entry code 3FAK), EstE5 (48.8 % sequence identity, PDB entry code 3FAK) and microbial E40[189] (49.4 % sequence identity, PDB entry code 4XVC). The two closest homologues (EstE5 and E40) were further employed to build a homology model with the free access RaptorX software.[166]

According to the sequence analysis the *Sv*Dea belongs to the HSL family, which is characterised by a hydrolytic mechanism based on a catalytic triad formed by conserved Ser, His and a Glu/Asp residues. As shown in Figure 4.11, H267 and E237 are highly conserved between the homologous, along with the amino acid motif GXSXG (from position 141 to 145), where X could by any amino acid. This motif usually contains the nucleophilic serine (S143 for *Sv*Dea) that is essential for the metabolism.[189]

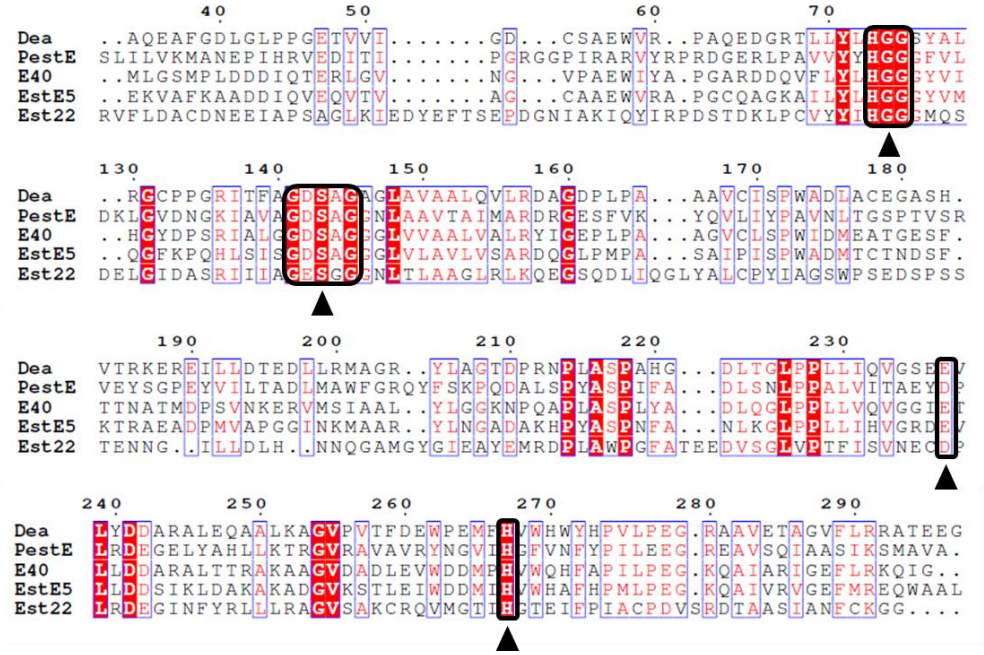


Figure 4.11 Sequence analysis of *Sv*Dea with other members of the HSL family. Sequence alignment between *S. viridochromogenes* Dea, Est22 (27.4 % sequence identity, PDB entry code 5HC0), *Pyrobaculum calidifontis* PestE (33.8 % sequence identity, PDB entry code 3FAK), EstE5 (48.8 % sequence identity, PDB entry code 3FAK) and microbial E40 (49.4 % sequence identity, PDB entry code 4XVC). White over red residues are conserved across all species, red on white are conserved across most while black on white are variable across species. The residues circled in black and underlined by a triangle are involved in the formation of the catalytic triad (S143, E237 and H267) and oxyanion hole (H73, G75 and G75).

The homology model shows the typical structure for a member of the HSL family, with a monomer composed of a CAP domain, formed by three α-helixes connected by loops (in orange in Figure 4.12 A) and a catalytic domain with a α/β-hydrolase fold, formed by eight β-strands and five α-helixes (in blue in Figure 4.12 A). This domain contains both the active site and the oxyanion hole, responsible for the stabilization of the reaction intermediate, which is usually formed by the HGG motif, His73-Gly74-Gly75 (Figure 4.11), in close proximity of the catalytic triad (Figure 4.12 B).

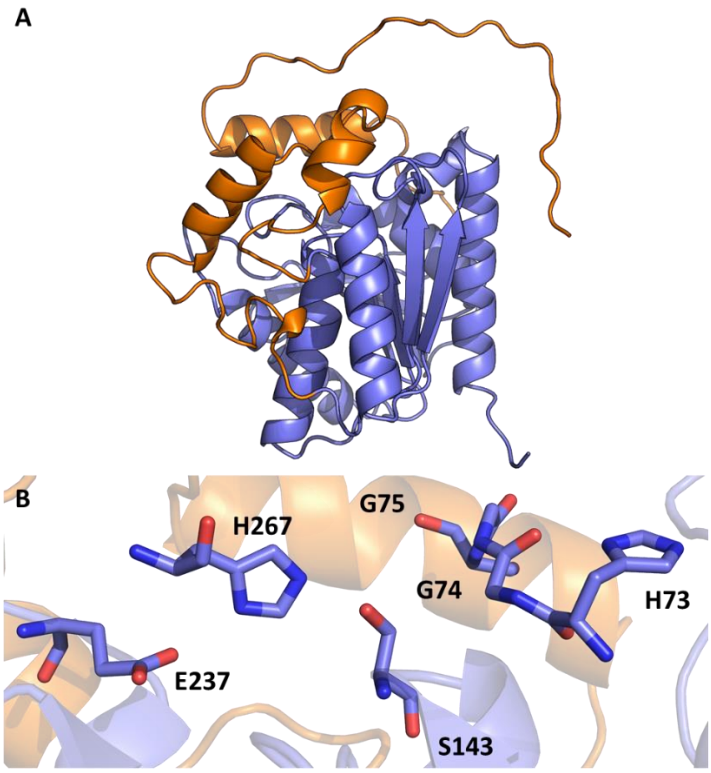
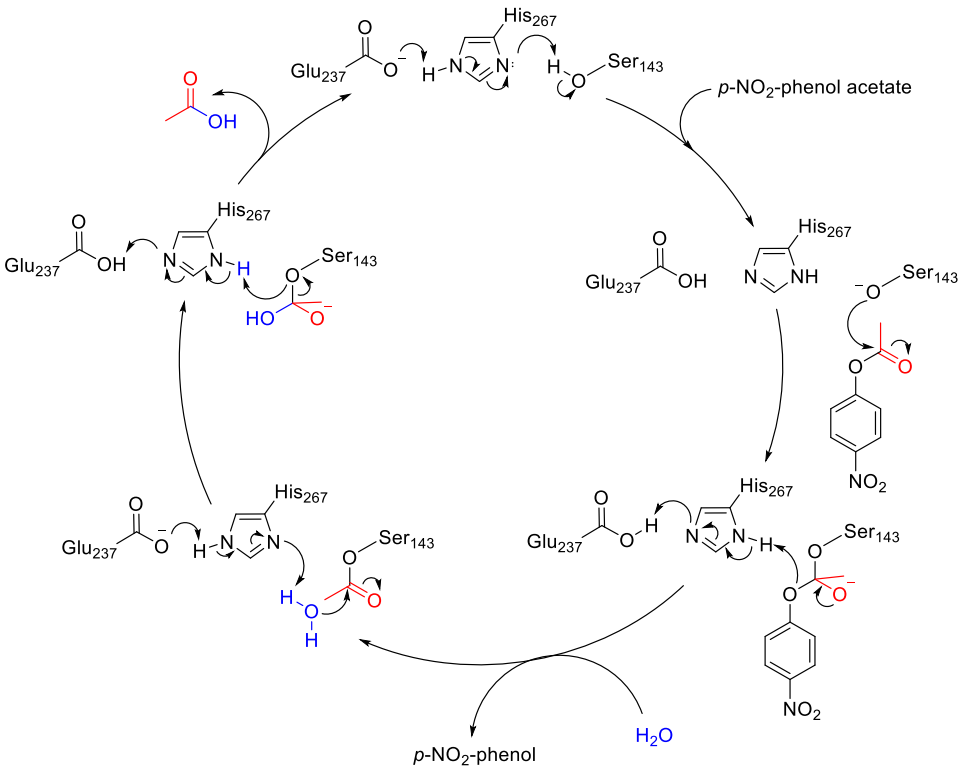


Figure 4.12 *Sv*Dea homology model built with the free access software RaptorX, based on the crystal structure of: bacterial EstE5 (3FAK) and microbial esterase E40 (4XVC). **A)** Tertiary structure of *Sv*Dea monomer; in blue the catalytic domain with an α/β-hydrolase fold and in orange the CAP domain. **B)** *Sv*Dea active site identified by sequence alignment; the E237, H267 and S143 residues form the catalytic triad, while the conserved H73, G74 and G75 motif generates the oxyanion hole.

There is a conserved mechanism for this family of hydrolytic enzymes (Scheme 4.5). It begins with the conserved Glu-His-Ser triad, which is connected by a network of H-bonds, increasing the nucleophilicity of the Ser143 hydroxyl group and facilitating the attack on the carboxylate group of the substrate. The negatively charged tetrahedral intermediate, stabilised by interaction with the oxyanion hole, collapses and releases the free alcohol or amine, leaving the catalytic Ser acylated. A water molecule, activated by H-bonding with the imidazole group of His267, attack the carboxylate, forming a second tetrahedral intermediate, which collapse releasing acetic acid and re-generating the catalytic triad.



Scheme 4.5 The *Sv*Dea catalytic triad five-step reaction cycle for the hydrolysis of *p*-nitrophenol acetate.

4.8 Conclusion and Future Work

In this work the novel *S. viridochromogenes* Tü494 deacetylase (*Sv*Dea) was expressed, purified and characterised. Activity studies carried out employing the continuous colorimetric *p*NP and L-AAO assays revealed that the enzyme possess a narrow substrate scope, being able to accept only esters with a short acyl chain (*p*NPA), while no activity was detected for the biotransformations of *N*-acetylated canonical or NCAAs, including *N*-Ac-PT and derivatives. Only traces of L-PTT formation were detected by LC-ESI MS, however these experiments need to be repeated on a larger scale to obtain conclusive results. These preliminary data reveal that *Sv*Dea does not possess biocatalytic potential for the synthesis of amino acids, but further studies are required to understand if the enzyme can accept other acetylated esters or derivatives as substrates.

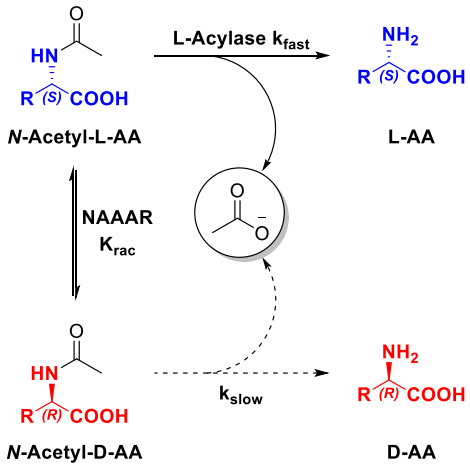
Initial crystal trials, performed using an N-hexahistidine tagged construct, failed to deliver any positive hits; more screening will be required in order to obtain protein crystals and more constructs (untagged and C-His tag) needs to be prepared and tested, the first step will be

the optimization of the TEV protease protocol. It is clear that *Sv*Dea is not stable overnight in the conditions used for dialyses, a shorter incubation time for digestion may be more effective and still deliver a good yield of non-tagged enzyme.

Sequence alignment and homology modelling revealed that *Sv*Dea belongs to the HSL family, has a structure formed by a capping and a catalytic α/β-hydrolase fold domains and uses a catalytic triad formed by the highly conserved S143, E237 and H267 residues to catalyse the deacetylation reaction. The primary function of the HSLs is to regulate the catabolism of triglycerides during fasting in many organisms,[190] it is possible that *Sv*Dea carries out a similar function in *Streptomyces* since the overexpression of the enzyme in *E. coli* cells appears to affect the composition of the bacteria membrane and wall, giving the cells a viscous consistency, which prevent the formation of proper pellets during harvesting and hinder the preparation of cell free extracts during purification. Further studies are required to confirm this hypothesis, but it is possible that *Sv*Dea acts as a bifunctional enzyme, intervening in the fatty acids' metabolism and in the biosynthesis of L-PTT tripeptide in *S. viridochromogenes*.

5 The *N*-Acetyl Amino Acid Racemase (NAAAR)

The *Amycolatopsis* sp TS-160 NAAAR was isolated by Tokuyama *et al.* in 1995[88] from a soil sample (Osaka). The enzyme selectively catalyses the racemization of *N*-acylated amino acids, but showed no activity for the free amino form of either L- or D-AAs. NAAAR has a wide substrate scope and can accept a range of aliphatic and aromatic amino acids with different *N*-acyl moieties: formyl (Fr), acetyl (Ac), chloro-acetyl (Cl-Ac), propionyl (Pr), succinyl (Sc), butyryl (Bt) and benzoyl (Bz).[88,107] In addition to its favourable substrate scope, NAAAR also has high thermostability and a metal cation cofactor. Therefore, NAAAR has great potential coupled with an L- or D-acylase for the development of a dynamic kinetic resolution (DKR) for the production of L- or D-amino acids respectively (Scheme 5.1).[109]



Scheme 5.1 Schematic representation of a NAAAR/L-Acylase catalysed DKR for the production of enantiopure L-AA with a theorical 100% conversion and *ee*, starting from a racemic mixture of *N*-Ac-DL-AA.

However, there are two main limitations of the WT NAAAR enzyme for industrial application in biocatalysis: substrate inhibition occurs at concentration > 50 mM and the enzyme has a preference towards longer acyl moiety (*N*-propionyl and *N*-succinyl AAs),[88,107] which has limited its application for industrial scale synthesis. In an effort to obtain a NAAAR variant with higher racemase activity for *N*-acetylated substrates, the enzyme was subjected to rounds of directed evolution and mutagenesis coupled with a live/death selection system of methionine auxotrophic pressure, yielding the double mutant G291D F323Y (DM) with 6-fold higher catalytic activity towards *N*-Ac-Met than the WT.[121]

The live/death assay was limited to methionine, so a novel colorimetric assay[122] (Scheme 5.2, see below) was developed to be able to screen NAAAR activity towards a wide range of

substrates. After the development of a high-throughput assay[122] for efficient screening of mutant libraries, a second round of mutagenesis was performed. This new "engineering" strategy was also guided by structural studies (Table 5.1). This led to the isolation of the quadruple mutant NAAAR Q26A M50I G291D F323Y (QM) with a 40-fold higher activity for *N*-Ac-Phe compared to the WT (data not published).[191]

5.1 Aims

The aim of this project was to express, purify and characterize the three active NAAAR constructs: WT, DM and QM (Table 5.1), in order to develop an efficient biocatalyst for the production of enantiopure non-canonical amino acids (NCAAs). Additionally, the conditions for the DKR will require optimisation alongside a wider study of the substrate scope for the mutants.

NAAAR	mutations	Uniprot code	PDB code	Ligand bound
WT	-	Q44244	1SJC	*N*-Suc-Met
DM	G291D F323Y	Q44245	4A6G	*N*-Ac-Met
QM	Q26A M50I G291D F323Y	Q44247	5FJU	*N*-Ac-2-Nal

Table 5.1 List of NAAAR variants, used in this work, with their mutations, Uniprot code, PDB entry code and the ligand bound in the crystal structure. With 2-Nal being the abbreviation for 3-(2-Naphthyl)-alanine.

5.2 NAAAR WT and DM purification and characterization.

The codon optimised *A.sp* TS-160 *naaar* constructs: pET20b-NAAAR WT and pET22b-NAAAR DM, were expressed and purified using the optimised protocol reported by Sànchez-Carròn.[122] The two enzymes were expressed in *Escherichia coli* BL21 (DE3) cells in TB media with 0.4 mM IPTG at 20°C for 20 h. The stability of NAAAR construct is affected by the presence of hexa-histidine tags,[192] either at the *N*- or *C*-terminus, therefore, all the recombinant proteins expressed are un-tagged; therefore to obtain the enzymes with a good purity level a three-step purification protocol was employed. NAAAR was separated from the more thermolabile proteins by subjection of the cell lysate to a thermal treatment at 60°C for 30 min, followed by anion exchange chromatography on a Q Sepharose Fast Flow (QFF) column with the enzymes eluting at a concentration of 0.3 M NaCl (Figure 5.1 A). Finally, size

exclusion chromatography (SEC) was performed on a pre equilibrated S300 Sephacryl SEC column.

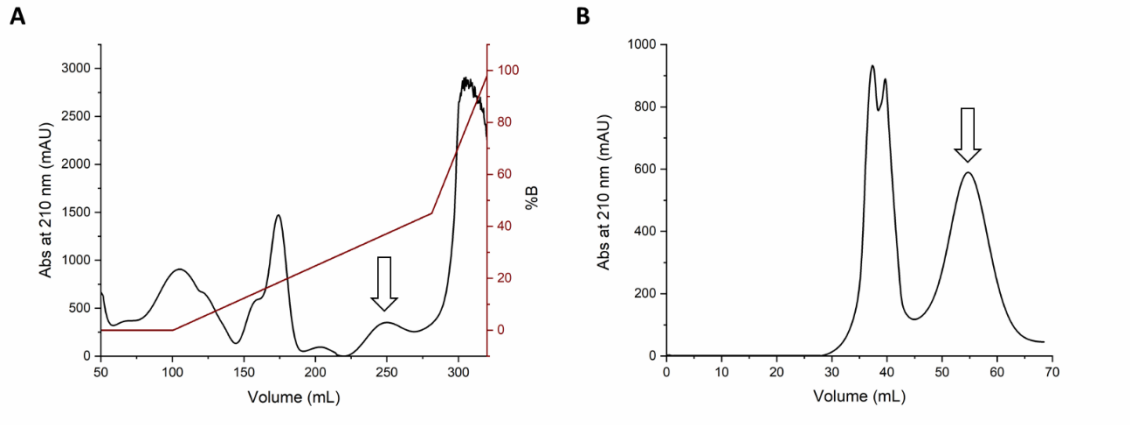


Figure 5.1 Purification of recombinant NAAAR DM. **A)** Anion exchange chromatogram for NAAAR DM purification, in black the UV-Vis traces and in red the elution profile expressed as gradient of %B. **B)** SEC chromatogram for NAAAR DM on a Sephacryl S300 column. The arrows point to the NAAAR elution peaks.

The UV/Vis chromatogram recorded at 280 nm displayed two symmetrical peaks, however analysis *via* SDS-PAGE (Figure 5.2 A and Figure 5.3 A) electrophoresis confirms that NAAAR (band at ~39 kDa) is present only in the second peak, at an elution volume of 55 mL (Figure 5.1 B), which corresponds to a molecular weight of ~300 kDa for native NAAAR (homo-octamer).

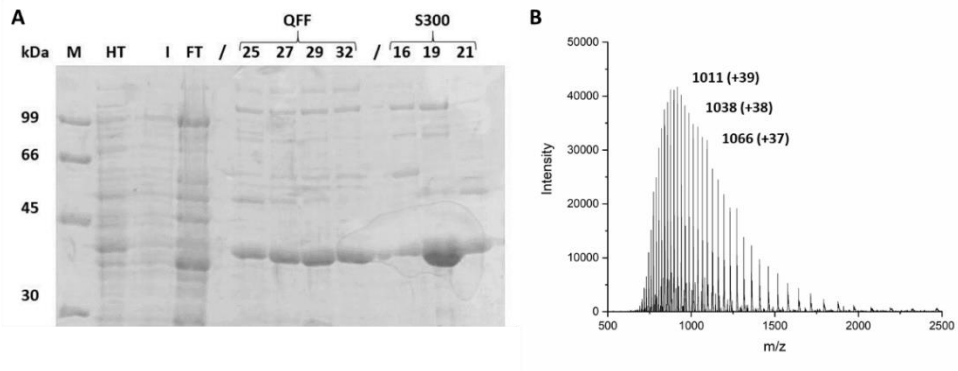


Figure 5.2 Characterization of NAAAR WT. **A)** 15% SDS-PAGE analysis of NAAAR WT purification: lane 1 LMWM, lane 2 CFE after heat-treatment, lane 3 insoluble fraction, lane 4 QFF flow-through, lane 5-8 QFF fractions (5 mL each) and lane 9-11 SEC fractions (3 mL each). **B)** Denaturating LC ESI-MS of NAAAR WT (20 µM) with a mass of 39406.02 ± 0.39 Da, in accordance with the MW calculated from the sequence (39406.42 Da). The values reported are the *m/z* with their respective charges (*z*) in brackets.

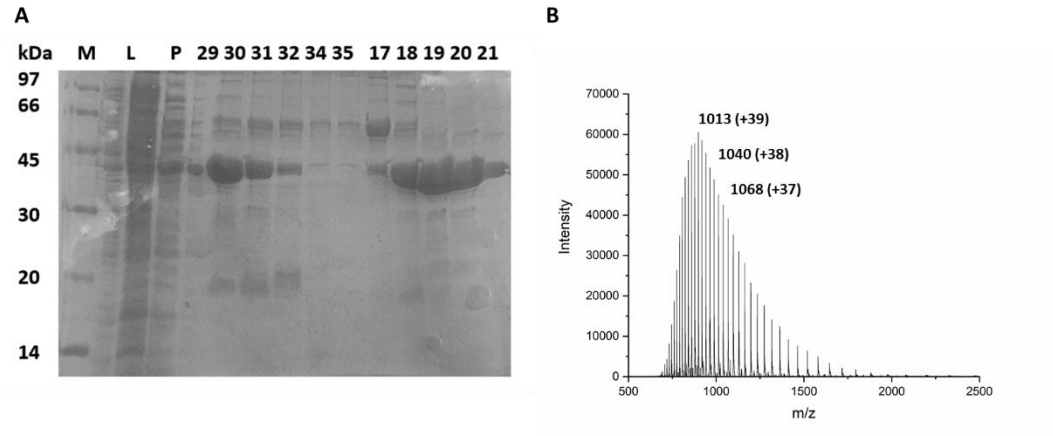
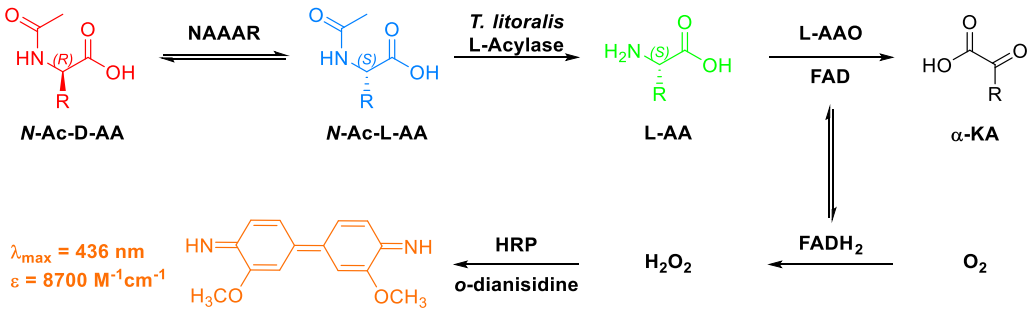


Figure 5.3 Characterization of NAAAR DM. **A)** 15% SDS-PAGE analysis of NAAAR DM purification: lane 1 the LMWM, lane 2 CFE after heat-treatment, lane 3 insoluble fraction, lane 4-9 QFF fractions (5 mL each, peak collected from ~225 to 275 mL) and lane 10-14 SEC fractions (3 mL each, ~45 to 65 mL). **B)** Denaturating LC ESI-MS of NAAAR DM (20 µM) with a mass of 39480.53 ± 0.41 Da, in accordance with the MW calculated from the sequence (39480.45 Da). The values reported are the *m/z* with their respective charges (*z*) in brackets.

After three purification steps NAAAR WT and DM were obtained with a purity of ~95% and an average yield of 25 mg *per* L of culture. To confirm that the correct constructs were expressed and purified, the molecular weight of NAAAR WT and DM were obtained using denaturing LC ESI-MS. The observed MW of 39406.02 ± 0.39 Da and 39480.53 0.41 Da, for WT and DM respectively (Figure 5.2 B and Figure 5.3 B), were in accordance with their theorical MW: 39406.42 Da for WT and 39480.45 Da for DM, calculated from the protein sequence with the ExPaSy ProtParam tool (https://web.expasy.org/protparam).

5.3 NAAAR WT and DM kinetic analysis

The high-throughput continuous L-AAO assay reported by Sànchez-Carròn[122] was employed for the kinetic analysis of NAAAR (Scheme 5.2). The *N*-Ac-D-AA substrate is converted by the racemase to the respective *N*-Ac-L-AA; which is hydrolysed by the *Thermococcus litoralis* L-Acylase (obtained from Dr. Reddy's Laboratories). The L-AA produced initiates a cascade of reactions, as described in Chapter 3.3.1, that results in the formation of an orange/red dye with a strong absorbance at 436 nm (ε = 8700 $M^{-1}cm^{-1}$).



Scheme 5.2 Reaction scheme of the NAAAR/L-AAO continuos assay.[122] The oxidised from of *o*-dianisidine has a strong absorbance at 436 nm and can be easily monitored *via* UV-Vis on a 96-well microplate reader.

The kinetic analysis was repeated in triplicates, at various substrate concentrations; the reactions were incubated at 50°C and the absorbance was recorded at 436 nm over 1 h on a microplate reader. As shown in Figure 5.4 A, the assay background is caused by the presence of the *T. litoralis* L-Acylase (a gift from Dr. Reddy's Laboratories)[134] in the reaction mixture. This enzyme was provided as lyophilized CFE and when dissolved in buffer is characterised by a faint orange colour. However, as the background absorbance is constant, the kinetic analysis is unaffected (Figure 5.4 B).

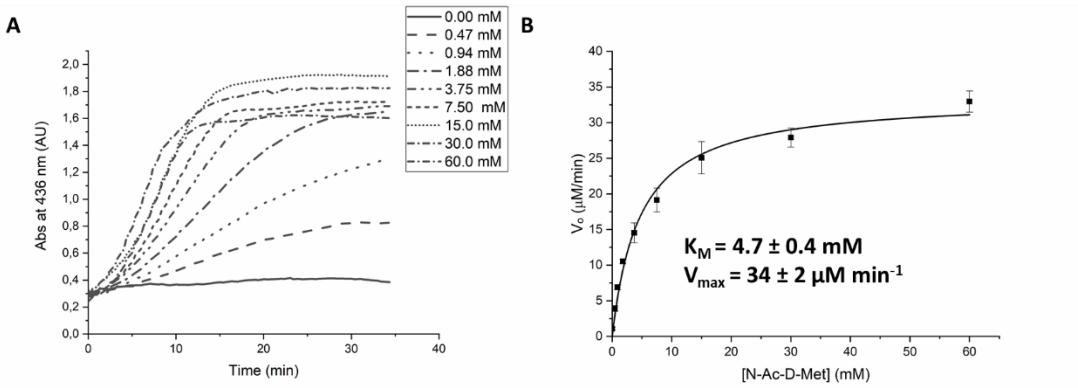


Figure 5.4 Kinetic analysis of NAAAR DM with L-AAO/HRP assay. **A)** Typical raw data profile for the formation of *o*-dianisidine$_{Ox}$ over time at various substrate (*N*-Ac-D-Met) concentrations**. B)** Michaelis-Menten plot of NAAAR DM for *N*-Ac-D-Met. Initial rates calculated from the linear portion of the absorbance curves and fitted with the OriginLab software. Reactions repeated in triplicates, error bars calculated as the standard deviations.

The kinetic parameters calculated using the DM (Table 5.2) are in accordance with the data reported in the literature[122] for *N*-Ac-D-Met, while for *N*-Ac-D-Phe a higher k_{cat} (0.46 ± 0.04 s^{-1}) and lower K_M (1.8 ± 0.5 mM) were obtained, compared to those reported by Sànchez Carròn: 0.104 ± 0.004 s^{-1} and 4.20 ± 0.39 mM.[122] However, this difference can be explained by the presence of $CoCl_2$ instead of $MgCl_2$ in the reaction mixture, since it is known that

NAAAR has a higher activity when Co^{2+} is used as the cofactor instead of Mg^{2+}.[193] Overall, as expected, the DM showed a higher activity for both substrates compared to the WT enzyme (Table 5.2 and Figure 5.5). Furthermore, given the high k_{cat} of NAAAR DM towards *N*-Ac-D-Met (0.87 ± 0.05 s^{-1}), this mutant was chosen for the coupling with the acylase ArgE.

NAAAR	WT		DM	
***N*-Ac-D**	**Met**	**Phe**	**Met**	**Phe**
K_M (mM)	3.7 ± 0.7	18 ± 3	4.7 ± 0.8	1.8 ± 0.5
k_{cat} (s^{-1})	0.095 ± 0.005	0.30 ± 0.04	0.87 ± 0.05	0.46 ± 0.04
k_{cat}/K_M ($M^{-1} s^{-1}$)	26.61 ± 0.01	17.07 ± 0.02	187.621 ± 0.005	243.03 ± 0.03

Table 5.2 NAAAR WT and DM kinetic parameters for *N*-Ac-D: Met and Phe, calculated with the L-AAO/HRP assay. Reactions repeated in triplicates, error bars calculated as the standard deviations.

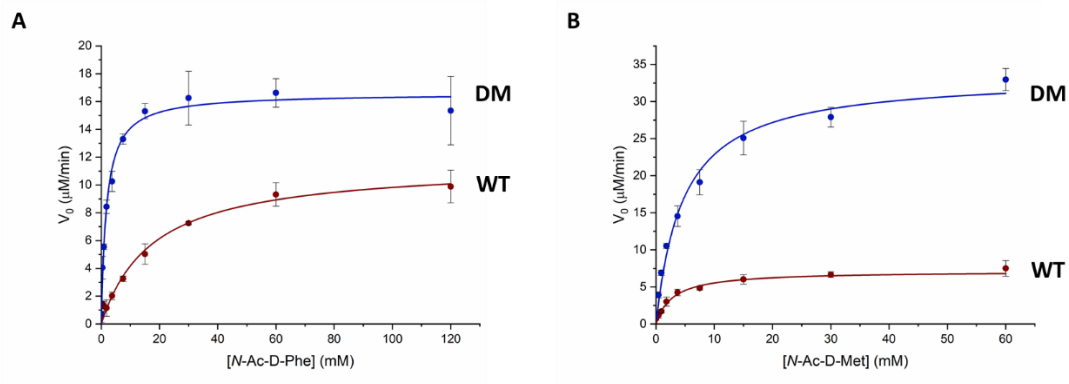
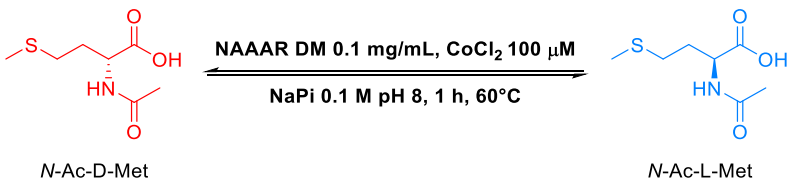


Figure 5.5 Kinetic curves of NAAAR WT and DM. Comparison between the Michaelis-Menten plots of NAAAR WT (red) and DM (blue) for **A)** *N*-Ac-D-Phe and **B)** *N*-Ac-D-Met. Reactions repeated in triplicates, error bars calculated as the standard deviations.

5.4 NAAAR DM Reaction Monitoring

The L-AAO assay provides a quick and efficient method to test NAAAR activity, nevertheless it relies on a system of coupled reactions. A more direct approach to monitor the racemization can be achieved *via* chiral HPLC, exploiting the amide bond absorbance at 210 nm.[141]



Scheme 5.3 Reaction scheme of the racemization of *N*-Ac-D-Met catalysed by NAAAR DM.

The racemization of *N*-Ac-D-Met (Scheme 5.2) was monitored over time using the same protocol described in Chapter 3.4 and 6.9.3. As shown in Figure 5.6, the racemization proceeds quite rapidly even at moderate enzyme loading (0.1 mg/mL) with a 42 % conversion of D- to L- in only 30 min; after one-hour incubation at 60°C, the enantiopure substrate was completely converted to a racemic mixture.

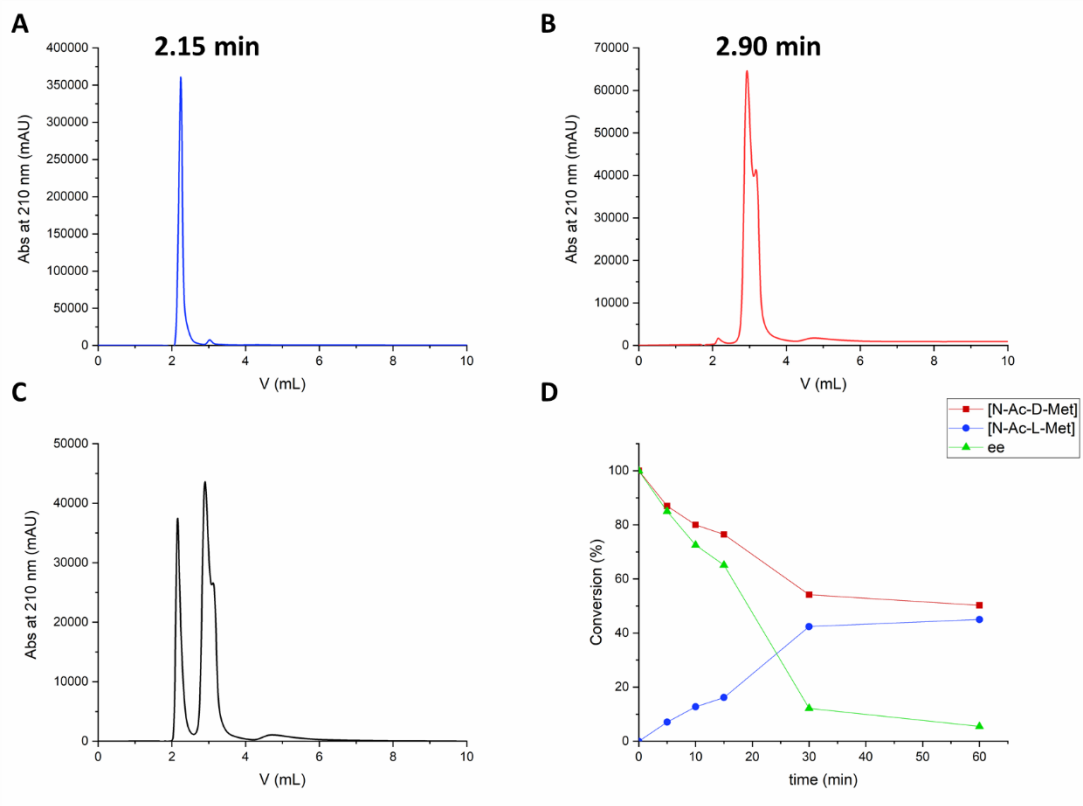


Figure 5.6 Chiral HPLC analysis of *N*-Ac-D-Met racemization. **A)** Elution profile of *N*-Ac-L-Met standard with a retention time (R.T.) of 2.15 min. **B)** Elution profile of *N*-Ac-D-Met standard with a R.T. of 2.90 min. **C)** Analysis of *N*-Ac-D-Met racemization sample following 30 min incubation at 60°C with NAAAR DM. **D)** Conversion of *N*-Ac-D-Met into *N*-Ac-L-Met catalysed by NAAAR. Time-point samples were analysed by Chirobiotic T (Teicoplanin) column with an isocratic gradient: elution phase 0.025% TEAA and MeOH (50/50 ratio), flow 1.0 mL/min, T = 40°C and λ = 210 nm. In red *N*-Ac-D-Met, in blue *N*-Ac-L-Met, in black the reaction mixture and in green triangles the solution *ee*. Reactions repeated in duplicates, error bars calculated as standard deviations reported only for triplicates.

5.5 Coupling of ArgE and NAAAR DM

5.5.1 DKR of *N*-Ac-DL-Met

The NAAAR DM and *E. coli* ArgE have the potential to be coupled together in an efficient DKR: they both have high TONs and work in cooperative conditions; active at pH 8 and with high $CoCl_2$ concentrations. However, a compromise had to be made on a reaction temperature of

50 degrees since ArgE (melting temperature of 65°C) is not a thermostable protein (Chapter 3.6) and, at the same time, a reaction temperature of 40°C is far from ideal for NAAAR activity.[88] To verify if ArgE still retained good activity at 50°C, an L-AAO assay was performed incubating various concentrations of NAAAR DM with 0.5 mg/mL ArgE, which replaced the *T. litoralis* L-Acylase in the assay mixture. As shown in Figure 5.7, ArgE is active and can couple efficiently with NAAAR DM at 50°C for the conversion of *N*-Ac-D-Met to L-Met. Furthermore, the biotransformation is very rapid, with the assay reaching saturation in less than 20 minutes at higher DM concentrations.

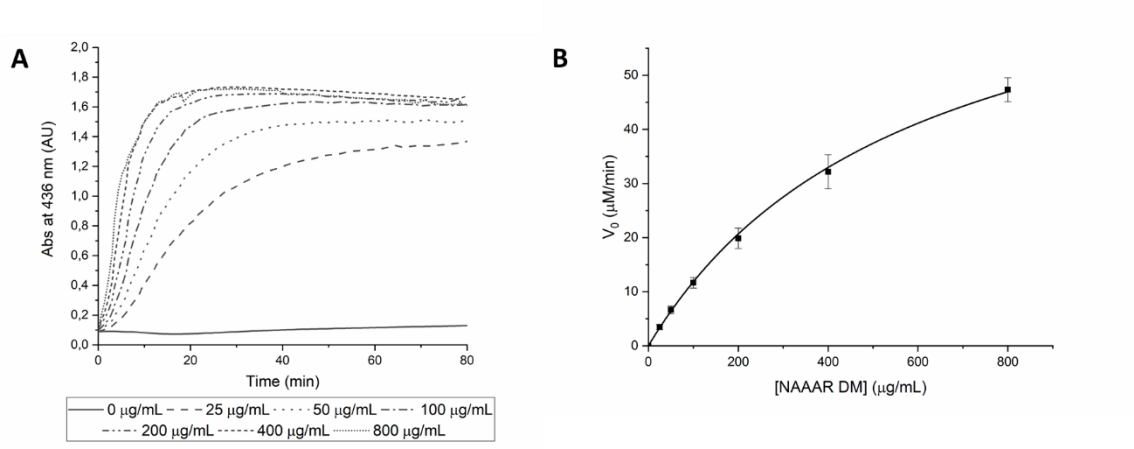
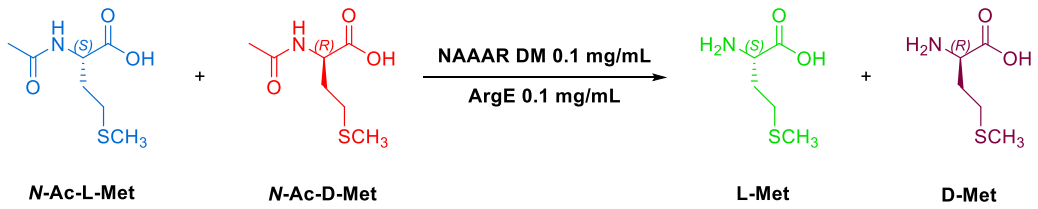


Figure 5.7 Coupling of NAAAR DM and ArgE monitored with the L-AAO assay. **A)** Raw data profile for the L-AAO assay, absorbance curves at 436 nm over time for the coupling reactions between ArgE 0.5 mg/mL and NAAAR 0.0-0.8 mg/mL with 15 mM *N*-Ac-D-Met at 50°C. **B)** Saturation curves between [NAAAR] and reactions initial rates calculated from the initial linear portion of the absorbance curves. Reactions repeated in triplicates, error bars calculated as the standard deviations.

Using these optimised conditions, a DKR of *N*-Ac-DL-Met (60 mM) was set up (Scheme 5.4). Reaction samples were taken over time, quenched by a 40-fold dilution into the chiral phase (75:25=0.025% TEAA:MeOH) and analysed by chiral HPLC.



Scheme 5.4 DKR of *N*-Ac-DL-Met 60 mM with 0.1 mg/mL ArgE, 0.1 mg/mL NAAAR DM and $CoCl_2$ 100 µM in NaPi 0.1 M pH 8. The reaction was incubated at 50°C for 24 h (250 rpm) and time points taken at: 0, 10, 20, 40, 60, 90, 120 and 180 min.

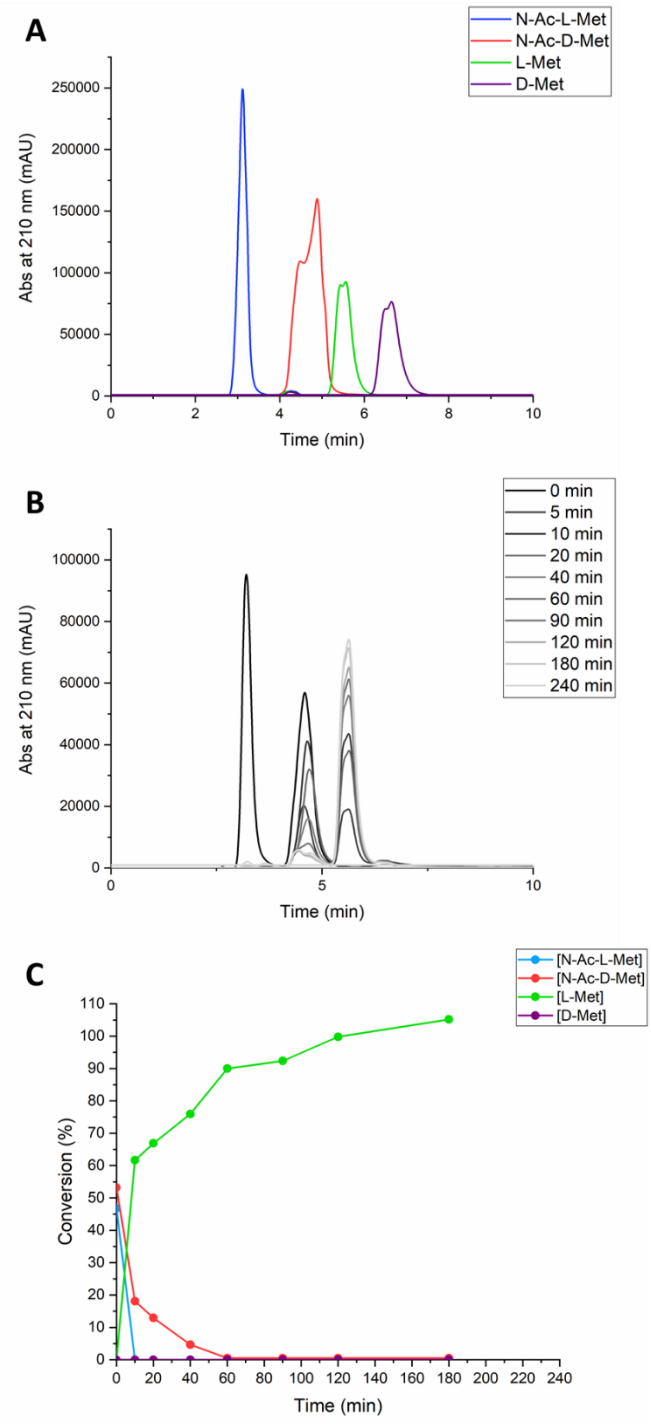


Figure 5.8 The DKR of *N*-Ac-DL-Met catalysed by NAAAR DM and ArgE. **A)** HPLC profile for the resolution of both enantiomers of the starting material: *N*-Ac-L-Met (blue), *N*-Ac-D-Met (red) and product: L-Met (green), D-Met (purple). Standards samples at a concentration of 25 mM. **B)** HPLC curves of time point samples of the biocatalytic DKR of *N*-acetyl methionine (scale of grey). **C)** Conversion of *N*-Ac-L-Met (blue) and *N*-Ac-D-Met (red) into L-Met (green) over time, no trace formation of D-Met (purple) detected. All samples analysed by HPLC on a Chirobiotic T (Teicoplanin) column, isocratic gradient: elution phase 0.025% TEAA and MeOH (75/25 ratio), flow 0.5 mL/min, T = 40°C and λ = 210 nm. Reactions repeated in duplicates, error bars calculated as standard deviations reported only for triplicates.

As shown in Figure 5.8 the hydrolysis is very rapid, after 10 minutes incubation, all *N*-Ac-L-Met starting material has been converted to L-Met. The racemization proceeds more gradually, but as soon as *N*-Ac-D-Met is converted by the DM into the L-stereoisomer, ArgE quickly hydrolyses it to form the final product, with the all process reaching completion in three hours. A faster DKR can be achieved by increasing the concentration of the two enzymes in the reaction mixture, however to effectively monitor the process by HPLC the DM and ArgE loadings have been purposely kept low. After 24 h incubation, the remaining reaction mixture was quenched by addition of HCl_{conc} and analysed by ^{1}H-NMR spectroscopy, which confirmed complete conversion of the substrate to L-Met.

5.5.2 Comparison between *E. coli* ArgE and *T. litoralis* L-Acylase

A second DKR of *N*-Ac-DL-Met was set up, which involved coupling NAAAR DM with the *T. litoralis* L-Acylase from Dr. Reddy's. Analysis by chiral HPLC again shows a fast hydrolysis and a gradual racemization with complete conversion achieved after 3 h incubation (Figure 5.9). Nevertheless, the process is slower compared to the DKR of *N*-Ac-methionine catalysed by NAAAR DM and ArgE. After the initial 10 minutes *N*-Ac-L-Met was still present in the reaction mixture and following 1 h incubation a conversion of just 77% was achieved, opposed to the 90% conversion obtained with ArgE; an impressive result considering that the ArgE concentration in the reaction mixture (0.1 mg/mL) was much lower than Dr. Reddy's L-Acylase (1 mg/mL lyophilized CFE). To note that no D-Met was formed during the reaction.

In conclusion *T. litoralis* L-Acylase has great potential application for the resolution of aromatic *substrates*, *N*-Ac: Phe, Tyr and Trp.[134] While with *E. coli* ArgE it is possible to achieve a rapid and efficient resolution of aliphatic and charged *N*-Ac-NCAAs. Between the *T. litoralis* L-Acylase and the *E. coli* ArgE, it is possible to set up DKRs to cover a broad substrate profile, ranging from bulky hydrophobic to small and polar amino acids and NCAAs.

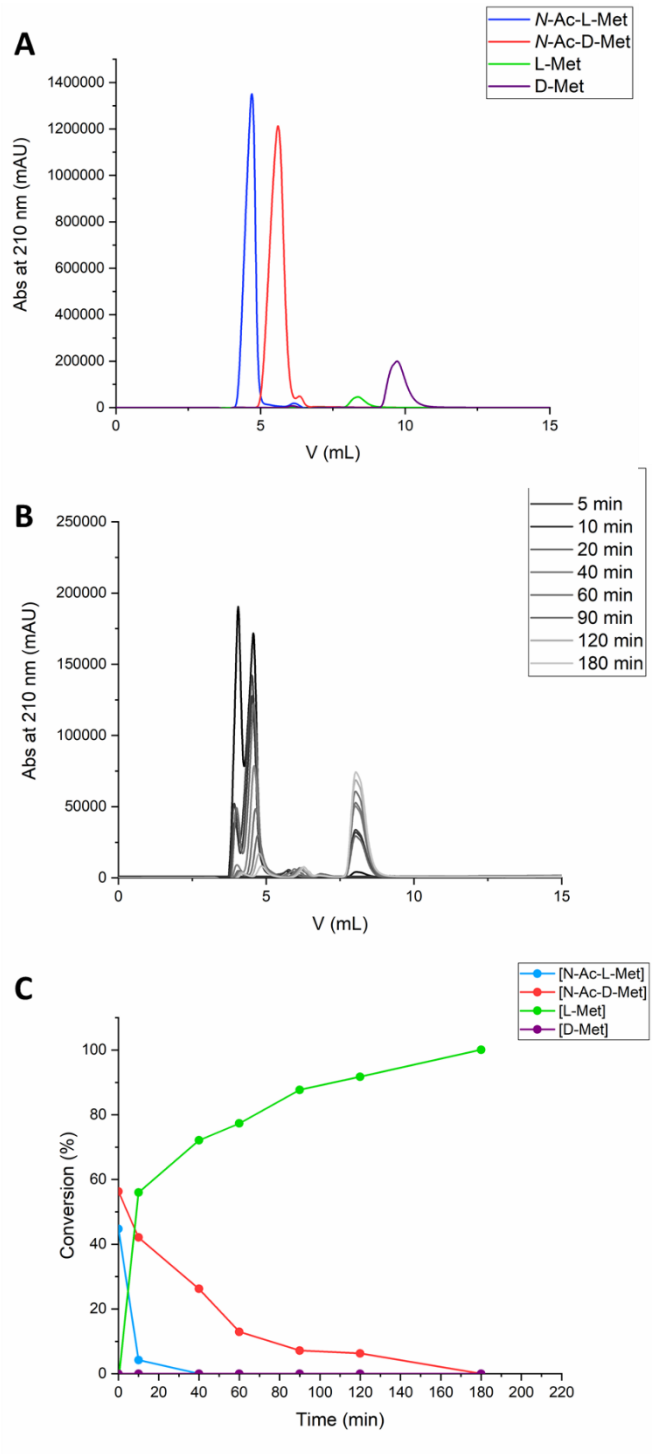
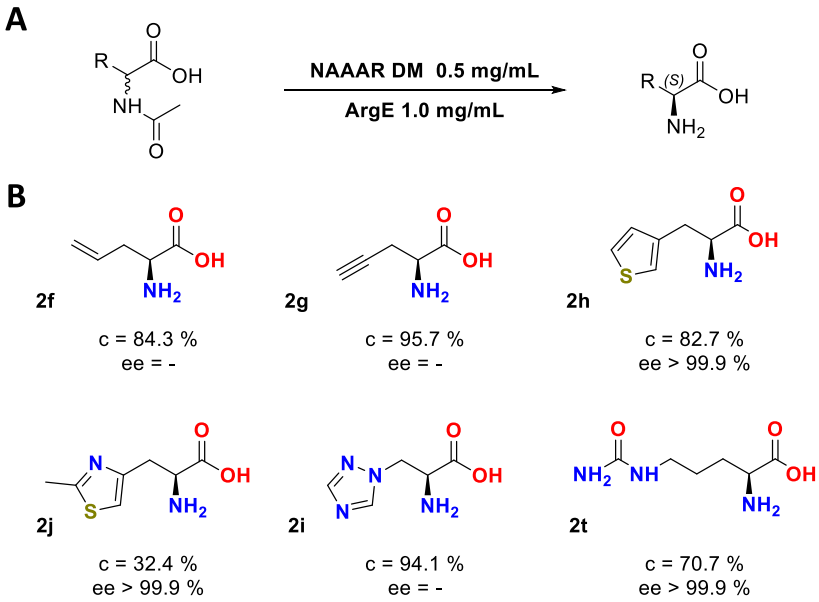


Figure 5.9 The DKR of *N*-Ac-DL-Met catalysed by NAAAR DM and *T. litoralis* L-Acylase. **A)** HPLC profile for the resolution of both enantiomers of the starting material: *N*-Ac-L-Met (blue), *N*-Ac-D-Met (red) and product: L-Met (green), D-Met (purple). Standards samples at a concentration of 25 mM. **B)** HPLC curves of time point samples of the biocatalytic DKR of *N*-acetyl methionine (scale of grey). **C)** Conversion of *N*-Ac-L-Met (blue) and *N*-Ac-D-Met (red) into L-Met (green) over time, no trace formation of D-Met (purple) detected. All samples analysed by HPLC on a Chirobiotic T (Teicoplanin) column, isocratic gradient: elution phase 0.025% TEAA and MeOH (40/60 ratio), flow 0.5 mL/min, T = 40°C and λ = 210 nm. Reactions repeated in duplicates, error bars calculated as standard deviations reported only for triplicates.

5.5.3 DKR of *N*-Ac-Non-Canonical-AAs

Given the promising results obtained in the DKR of *N*-Ac-Met, the NAAAR DM/ArgE couple was also employed for the resolution of *N*-Ac-NCAAs. It is known that NAAAR DM is able to accept a wide range of substrates, including some non-proteinogenic *N*-Ac-AAs (Scheme 5.5), such as: *N*-Ac-allylglycine (Alg), *N*-Ac-phenylglycine (*N*-Ac-Phg), *N*-Ac-4-F-Phg, *N*-Ac-4-NO_2-Phe and *N*-Ac-2-naphthylalanine (*N*-Ac-2-Nal).[122,192,194] The activity of NAAAR with the new *N*-Ac-NCAAs was also investigated. However, the enzyme loading for these experiments was increased from 0.5 mg/mL for the DM and ArgE to 1 mg/mL (the optimised enzyme concentration reported in Chapter 3.9.4, Scheme 5.5 A). The reactions were incubated for 24 h at 50°C and subsequently quenched with HCl_{conc} and analysed by ^{1}H-NMR spectroscopy (see Chapter 6.9.1).



Scheme 5.5 Substrate profile of NAAAR DM/ArgE DKR. **A)** Schematic representation of a DKR of an *N*-Ac-AA 50 mM catalysed by NAAAR DM 0.5 mg/mL and ArgE 1.0 mg/mL with 200 µM $CoCl_2$ in NaPi 0.1 M pH 8. Reaction incubated at 50°C 250 rpm for 24 h. All DKRs were quenched by HCl_{conc} addition and analysed by NMR spectroscopy and chiral HPLC. **B)** List of DM/ArgE DKR products with their ^{1}H-NMR spectroscopy conversions and *ee*. Reactions repeated in duplicates, error bars calculated as standard deviations reported only for triplicates.

As shown in Scheme 5.5 B, average to excellent conversions were obtained; with the best values (c > 90%) calculated for *N*-Ac-propargylglycine (**1g**) and *N*-Ac-1-triazolylalanine (**1i**). However, *N*-Ac-citrulline (**1t**) gave only modest conversions despite ArgE previously showed excellent activity towards polar and positively charged *N*-Ac-AAs (Chapter 3.8 and Chapter 3.9), suggesting it is NAAAR, which does not favourably bind to this polar substrate. The worst

conversion (c = 32.4%), "as expected", was obtained for the reaction with *N*-Ac-(2-Me-thiazolyl)-4-alanine (**1h**): indeed, ArgE presented low activity for this bulky heterocyclic substrate. However, there was an increase of 10% conversion upon the previous KR (Chapter 3.9.4) to 32% for the DKR, indicates that the DM is able to bind to the substrate and the racemization equilibrium pushes the process towards the formation of L-2-Me-thiazolyl)-4-alanine **(2j)**. Samples of each reaction were also analysed by chiral HPLC (protocol reported in Chapter 6.9.3) and the *ee* percentages was calculated for the substrates with an absorbance at 210 nm.

Protein precipitation was observed in all reactions after overnight incubation at 50°C, therefore, at high concentration (1 mg/mL), ArgE is most probably unstable under these biotransformation conditions. To optimize the process, a second set of DKRs were run at 40°C and at this temperature no precipitation was observed after 24 h incubation; however, with the exception of *N*-Ac-3-thienylalanine (**1h** c = 89.4%), all other ^{1}H-NMR conversions gave lower values compared to the previous reaction set (Table 5.3). The decrease in temperature likely slowed the racemization, the limiting step of the process. Perhaps by increasing the reaction time and/or the NAAAR concentration it may be possible to obtain better results even at 40°C, however so far, the best results were obtained at a reaction temperature of 50°C (Figure 5.10).

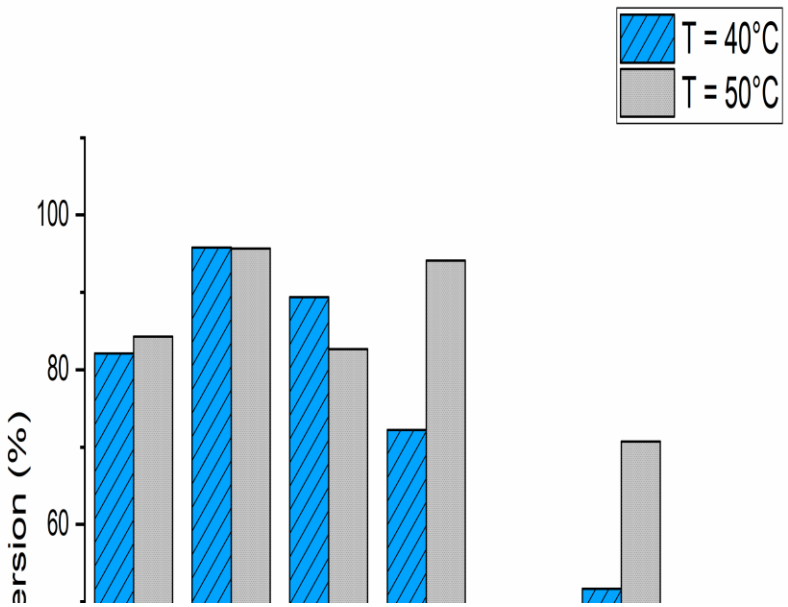


Figure 5.10 Comparison between the NAAAR DM/ArgE performances at 40°C (blue bar) and 50°C (grey bar) for the DKR of *N*-Ac-NCAAs. Conversions calculated by ^{1}H-NMR spectroscopy. Reactions repeated in duplicates, error bars calculated as standard deviations reported only for triplicates.

	^{1}H-NMR Conversion (%)	
T° =	**40°C**	**50°C**
1f	82.1	84.3
1g	95.8	95.7
1h	89.4	82.7
1i	72.2	94.1
1j	8.7	32.4
1t	51.7	70.7

Table 5.3 NAAAR DM and *E. coli* ArgE DKR conversions calculated by ^{1}H-NMR spectroscopy by monitoring the acetate peaks intensity for the reaction sets incubated at 40°C and 50°C.

5.6 NAAAR QM expression, purification and characterization

The NAAAR QM biocatalyst was expressed and purified using the standard NAAAR protocol (see Chapter 6.3 and 6.4).[122] The SEC chromatogram shows a peak at 60 mL (Figure 5.11 B), which according to the calibration curve, correspond to a 300 kDa protein; like the other variants, the QM also forms a homo-octamer in solution. SDS-PAGE analysis (Figure 5.11 A) confirms the presence of the protein, the gel presents a band at ~39kDa, approximately the MW of the monomer.

To further confirm the presence of the correct mutations, a sample of the purified enzyme was analysed *via* LC ESI-MS (Figure 5.11 C). The observed mass of 39409.38 ± 1.08 Da was in accordance with the theoretical MW (39405.37 Da) of the enzyme, calculated from the amino acid sequence. NAAAR QM was successfully purified with a yield of 20 mg *per* L of culture and a purity of 90%.

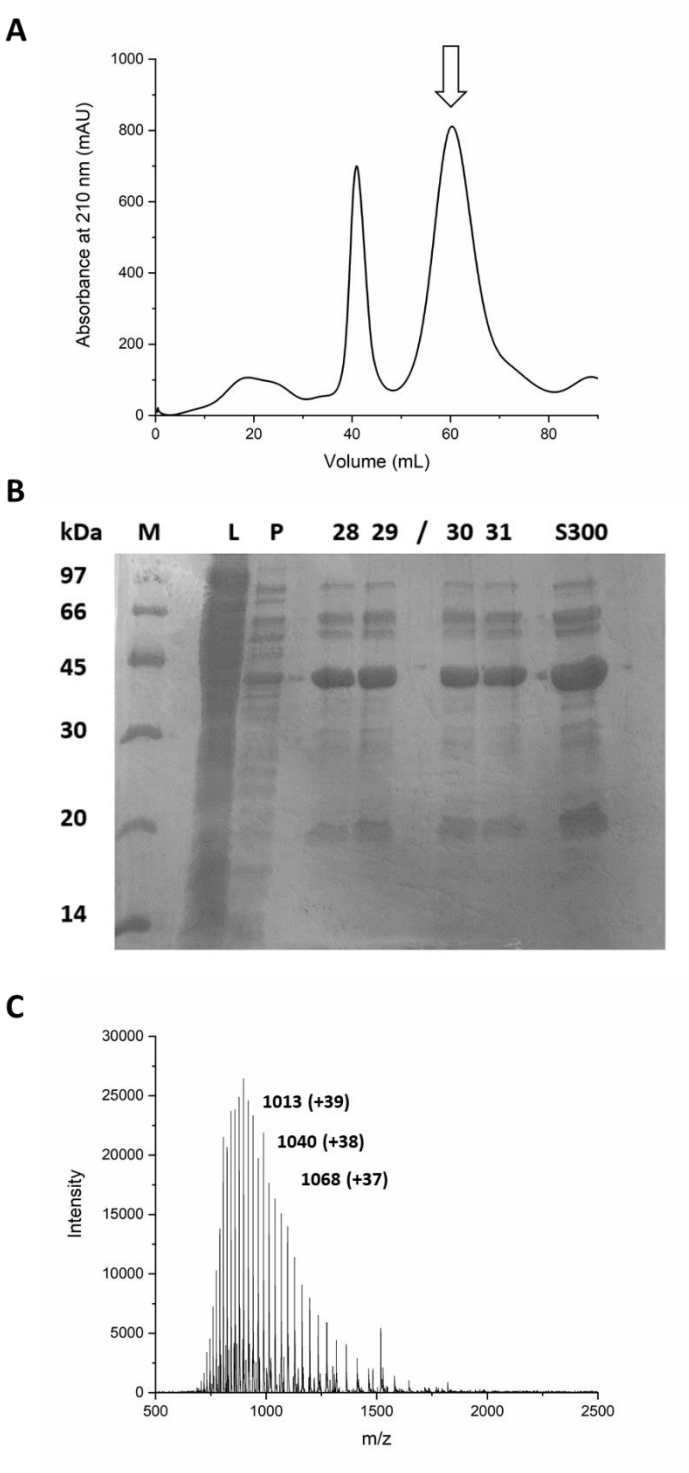
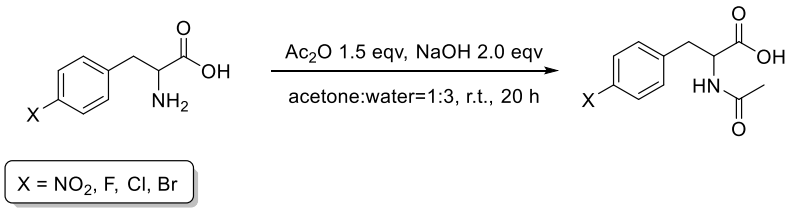


Figure 5.11 Purification and characterization of NAAAR QM. **A)** Size exclusion chromatogram for NAAAR QM on a Sephacryl S300 column. The arrow points to the NAAAR elution peak. **B)** 15% SDS-PAGE analysis of NAAAR QM purification: lane 1 LMWM, lane 2 CFE after heat-treatment, lane 3 insoluble fraction, lane 4-8 QFF fractions (5 mL each) and lane 9 S300 fraction (3 mL each). **C)** LC ESI-MS of NAAAR QM (20 µM) with a mass of 39409.38 ± 1.08 Da, in accordance with the MW calculated from the sequence (39405.37 Da). The values reported are *m/z* with the charges in brackets.

5.7 Investigating the NAAAR QM substrate scope

5.7.1 Synthesis of *N*-Ac-AAs substrates

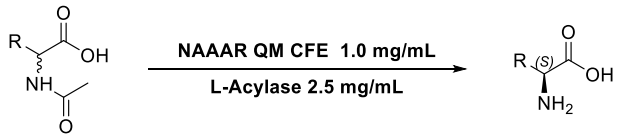
According to Sànchez-Carròn[191] the introduction of the mutations Q26A and M50I into NAAAR DM, opened the enzyme active site, generating an increase in catalytic activity but also a higher affinity for aromatic substrates. To test the potential of the novel QM as biocatalyst, a range of 17 racemic *N*-Ac-AAs were selected to use in a DKR process with the NAAAR QM. However, not all substrates were commercially available; *N*-Ac: Phe, Tyr, Trp and Phg were purchased from Sigma Aldrich, *N*-Ac: 1-triazolylalanine, 3-F-Phe and 2-F-Phe were obtained from Syngenta, *N*-Ac-2,2-bypyridylalanine (BpyA) was a gift from Dr Richard Brewster (University of Edinburgh, School of Chemistry) and racemic *N*-Ac: citrulline (**1t**), allylglycine (**1f**), propargylglycine (**1g**), 3-thienylalanine (**1h**) and 2-Me-thiazolyl-4-alanine (**1j**) were prepared as described in Chapter 3.9.1. *N*-Ac: 4-NO_2-Phe (**1u**), 4-F-Phe (**1v**), 4-Cl-Phe (**10a**) and 4-Br-Phe (**10b**) were prepared from their respective AA racemic mixtures by treatment with slight excess of acetic anhydride in presence of a strong base (Scheme 5.6). The products were isolated and characterised by NMR spectroscopy and MS (see Appendices).



Scheme 5.6 Scheme for the acetylation reaction of phenylalanine analogues. The amino acid solution (0.1 M in 1:3=acetone:water) was reacted with Ac_2O (1.5 eqv) and NaOH (2.0 eqv) for 20 h at r.t.

5.7.2 Small scale screening

A set of small scale (500 µL) DKRs was set up coupling NAAAR QM with *T. litoralis* L-Acylase (Scheme 5.7); interestingly precipitate formation was observed in almost all samples during the incubation. After isolating the solid by centrifugation; analysis with ESI-TOF MS revealed the precipitate to be the product which is insoluble in the reaction media. Potentially, the precipitation could also drive the equilibrium, leading to a further increase in the percentage conversions



Scheme 5.7 Small scale DKR of *N*-Ac-AAs 50 mM with NAAAR QM CFE (1.0 mg/mL) and *T. litoralis* L-Acylase (2.5 mg/mL with $CoCl_2$ 1 mM in NaPi 0.1 M pH 8. The reactions were incubated at 60°C, 250 rpm for 24 h and quenched by addition of NaOH 5M. Analysis carried using ^{1}H NMR and chiral HPLC.

After 24 h incubation at 60°C, the DKR conversions were obtained by quenching the reaction mixture by addition 20 µL of NaOH (5M). The product was redissolved and the crude reaction mixture was analysed by NMR spectroscopy to determine conversion (see Chapter 6.9.1) and chiral HPLC (see Chapter 6.9.3) for the *ee* of the UV-active products (Figure 5.12).

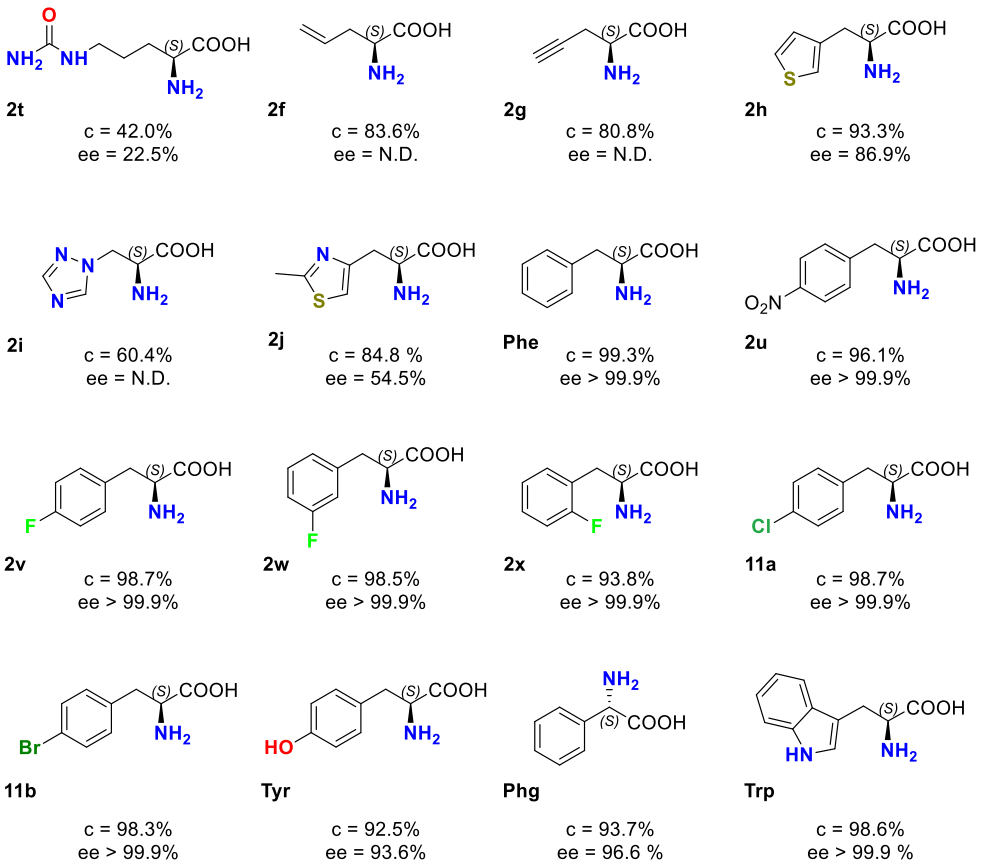


Figure 5.12 Summary of the of QM/ *T. litoralis* L-Acylase DKR substrate range (using sixteen *N*-Ac-AAs). Reaction products reported with their ^{1}H-NMR percentage conversions and *ee*. Reactions repeated in duplicates, error bars calculated as standard deviations reported only for triplicates.

In order to test the efficiency of the process for the production of both L- and D-AAs, another set of small scale (500 µL) DKRs were set up coupling the NAAAR QM with *Achromobacter xylosoxidans* subsp. *denitrificans* ATCC 15173 D-Acylase (obtained as lyophilized CFE from Dr. Reddy's Laboratories).[195] As with the *T. litoralis* L-Acylase reactions, the same protocol was

employed for the DKRs with *A. xylosoxidans* D-acylase, changing only the incubation temperature from 60° to 37°C, the ideal temperature reported for this D-Acylase.[195] As described in Section 5.5.3, a decrease of NAAAR performance has to be expected at T < 50°C. However, the high enzyme loading and longer incubation time will hopefully reverse any previously observed loss in activity.

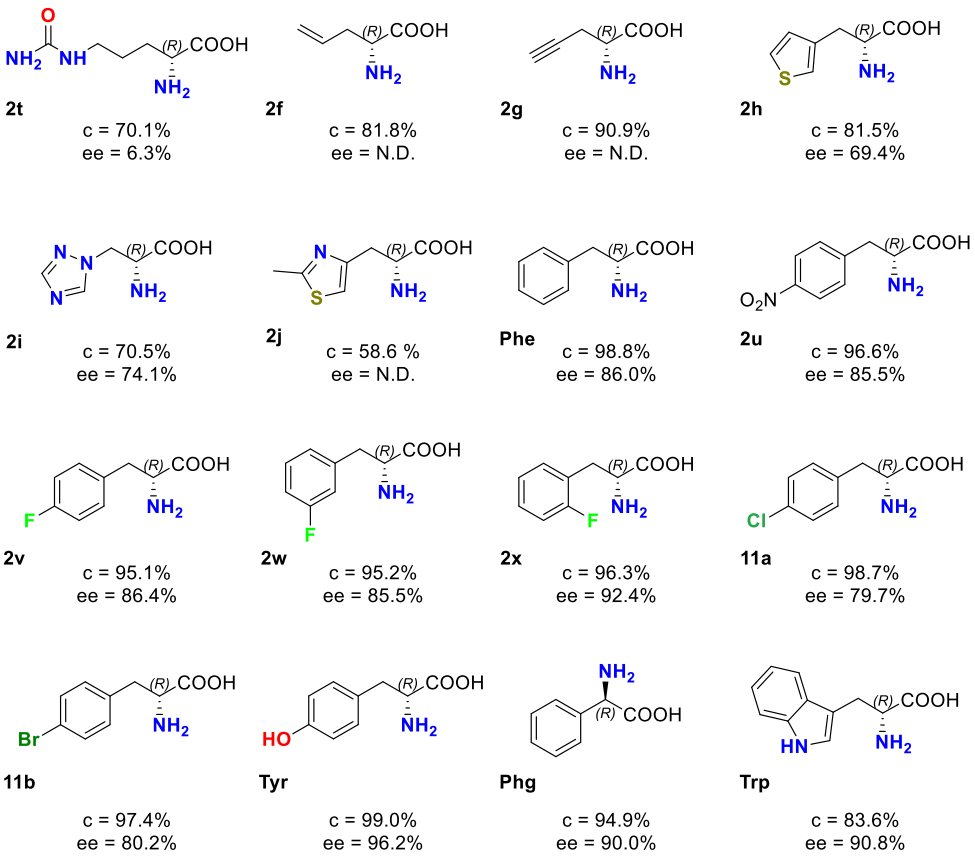


Figure 5.13 Summary of QM/ *A. xylosoxidans* D-Acylase DKR substrate range (using sixteen *N*-Ac-AAs). Reaction products reported with their ^{1}H-NMR percentage conversions and *ee*. Reactions repeated in duplicates, error bars calculated as standard deviations reported only for triplicates.

Across the sixteen substrates tested, good to excellent conversions were obtained for most substrates, using either the *T. litoralis* L-Acylase (Figure 5.12) or *A. xylosoxidans* D-Acylase (Figure 5.13) as the coupling enzyme, with the exception of **1t** which gave the lowest c and *ee* values. This suggests the NAAAR QM is not particularly active towards this polar amino acid; indeed, similar results were obtained for this substrate by coupling the DM with ArgE in Section 5.3.3.

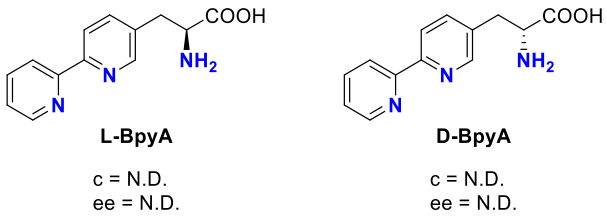


Figure 5.14 Structure of L and D-BpyA, no traces of either enantiomers detected from the DKR reaction mixtures after analysis by NMR, HPLC or TOF-ESI MS.

No activity was detected either by NMR spetroscopy, HPLC, or TOF-ESI MS for the resolution of *N*-Ac-BpyA (Figure 5.14), no traces of product detected by either: NMR, HPLC and TOF-ESI MS. Possibly the larger sterically hindered substrate cannot access the catalytic site of the enzymes or it may bind but not in the right conformation for the catalysis to occur. A similar case was reported by Sànchez-Carròn[191] for *N*-Ac-2-Nal; NAAAR QM showed poor activity towards this bulky amino acid, but the substrate was still able to efficiently bind to the mutant, indeed a ligand bound X-ray structure of QM in complex with *N*-Ac-2-Nal (PDB code 5FJU) was resolved.

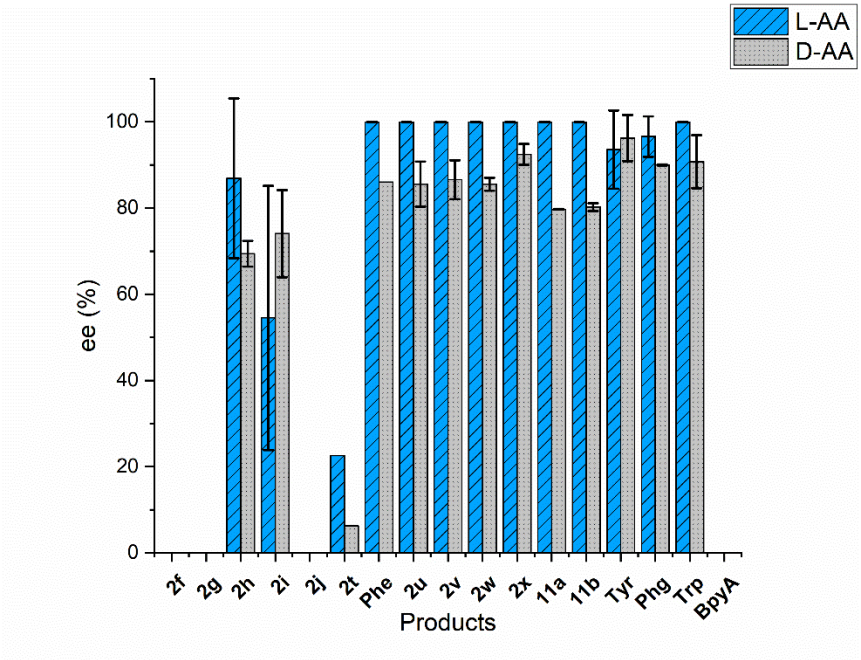
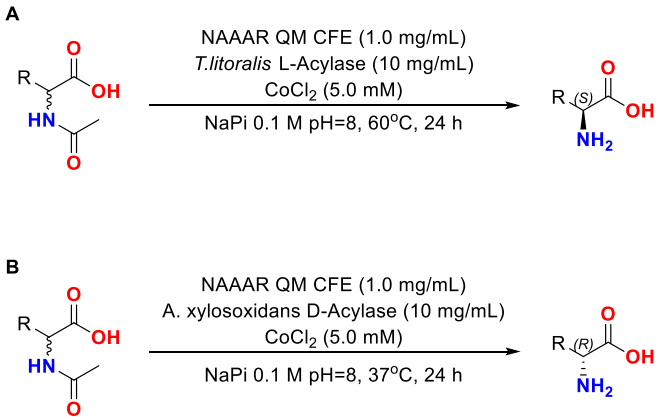


Figure 5.15 Comparison between the *ee* values obtained from the NAAAR QM DKR screening with the *T. litoralis* L-Acylase (blue bar) and *A. xylosoxidans* D-Acylase (grey bay). No results reported for **2f**, **2g** and **2j**, because the low UV-Vis absorbance of these AAs prevented analysis by chiral HPLC.

In general, better selectivity (*ee* > 90%) was obtained using the *T.litoralis* L-Acylase coupling enzyme than the *A. xylosoxidans* deacetylase (Figure 5.15). However, this factor is not indicative of a poor NAAAR QM performance, as this enzyme simply catalyses the

racemization equilibrium of *N*-Ac-AAs, but it is correlated with the enantioselectivity of the acylase used in the process.

5.8 DKR Scale-Up



Scheme 5.8 Preparative-scale DKR of *N*-Ac-aromatic AAs (0.5-1.0 g) with: NAAAR QM CFE (1.0 mg/mL), Acylase (10 mg/mL) and $CoCl_2$ (5 mM) in NaPi 0.1 M pH 8. **A)** Scheme for L-AA resolution with the *T. litoralis* L-Acylase, reaction T=60°C. **B)** Scheme for D-AA resolution with the *A. xylosoxidans* D-Acylase, reaction T=37°C. All DKRs were incubate at 250 rpm for 24h.

To prove that the NAAAR QM/Acylase couple can be used for the industrial-scale production of enantiopure amino acids, the DKR of a few *N*-Ac-aromatic AAs was scaled up (0.5-1.0 g substrate) and each reaction was repeated using both the *A. xylosoxidans* D-Acylase and *T. litoralis* L-Acylase as shown in Scheme 5.8. The DKR at scale were set up with higher substrate loading, 150-200 mM compared to 60 mM at analytical scale, the enzyme loading of the acylases was also increased to 10 mg/mL to avoid possible substrate inhibition.[191] All biotransformations were set up and incubated at 37 or 60°C, depending on the acylase used, for 24 h. After one day incubation a conversion of substrate to product > 90% was calculated by crude ^{1}H-NMR spectroscopy. The products were isolated by filtration and full chemical characterisation was carried out.

As shown in Figure 5.16, all yields obtained are quite high (> 60%), only exception is L-Phg resolution with a yield of 50.2%. Once more, the *T. litoralis* L-Acylase/NAAAR QM couple gave products with higher enantiopurity (>99% by chiral HPLC), while the *A. xylosoxidans* D-Acylase showed low to average stereoselectivity (Figure 5.17). The increase of the acylase loading in the scale-up process was counterproductive for the D-AAs resolutions, with lower

enantiomeric excess values, especially for the production of D-4-NO_2-Phe (**D-2u**) and D-4-F-Phe (**D-2v**), than those obtained in the small-scale screening (Chapter 5.7.2).

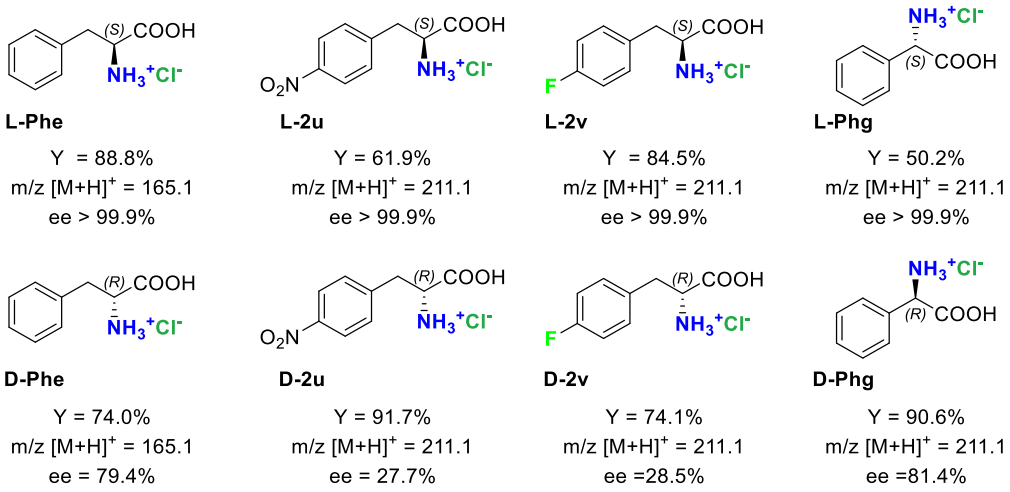


Figure 5.16 Isolated products obtained from the NAAAR QM DKR scale up, on the top row the L-AAs and on the bottom row the D-AAs, each reported with its percentage Yield, m/z and *ee*.

The performance of NAAAR QM for the resolution of *N*-Ac-AAs at gram-scale is comparable with the NAAAR DM DKR of *N*-Ac-allylglycine reported by Baxter *et al*,[121] who were able to isolate D-allylglycine with 89% yield and >99% *ee* after 18 h incubation. To note that to achieve those results with the DM an *E. coli* fermentation was necessary to produce enough enzyme, while a normal 2 L cell culture grow-up/expression, was sufficient to produce enough QM CFE for 8 scale up reactions. Given NAAAR QM wide substrate scope and the low enzyme loading required to obtain an efficient process, this racemase mutant has high potential for the synthesis of non-natural amino acids at industrial scale.

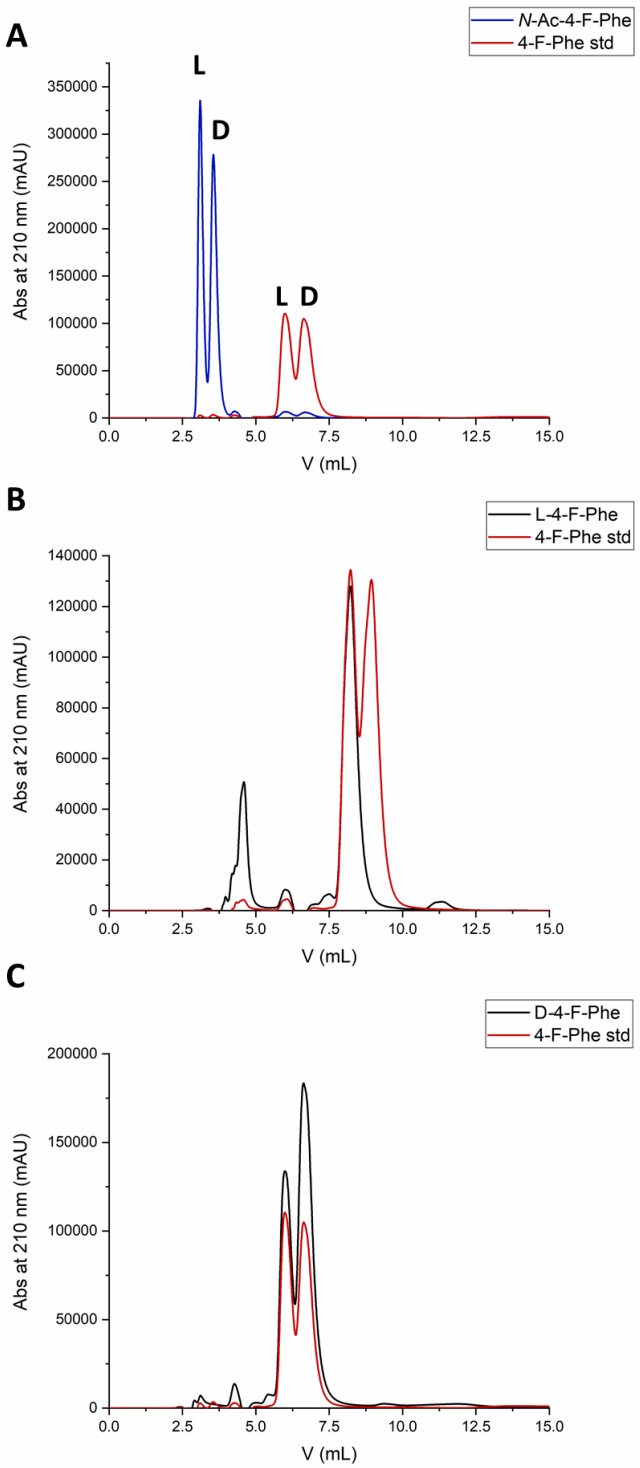


Figure 5.17 Chiral HPLC analysis of the preparative *N*-Ac-4-F-Phe (**1v**) DKRs. **A)** HPLC profile for the resolution of both enantiomers of the starting material *N*-Ac-4-F-Phe (blue) and product 4-F-Phe (red). Standards samples at a concentration of 25 mM. **B)** Superimposition of racemic 4-F-Phe standard (red) and L-4-F-Phe (**L-2v**) DKR sample (black) chiral HPLC curves. **C)** Superimposition of racemic 4-F-Phe standard (red) and D-4-F-Phe (**D-2v**) DKR sample (black) chiral HPLC curves. All samples analysed by HPLC on a Chirobiotic T (Teicoplanin) column, isocratic gradient: elution phase 0.025% TEAA and MeOH (40/60 ratio), flow 0.75 mL/min, T = 40°C and λ = 210 nm; reaction repeated in triplicate.

5.9 Conclusions and Future Work

The initial aims of the projects have been met. The three *Amycolaptopsis* sp TS-160 NAAAR variants: WT, DM (G291D F323Y) and QM (Q26A M50I G291D F323Y), have been fully characterised and the active mutants applied for the production of non-canonical amino acids.

The main target of the thesis was the biocatalytic synthesis of L-phosphinothricin (PT) *via* DKR; considering the aliphatic and charged nature of this amino acid side chain, NAAAR WT and DM were deemed to be the best choice to couple with ArgE, since the QM was engineered to accept aromatic *N*-Ac-AAs and ArgE is almost completely inactive towards these substrates (see Chapter 3.3 and 3.8).

The kinetic parameters of NAAAR WT and DM for two substrates (*N*-Ac-D-Met and *N*-Ac-D-Phe) were calculated, using the continuous coupled colorimetric L-AAO assay, and by comparing the WT and DM results, a 6-fold increase in activity (k_{cat}) was calculated for the reactions catalysed by the double mutant, in accordance with the results reported in literature.[121] Indeed, this NAAAR mutant was able to completely racemase a sample of enantiopure *N*-Ac-D-Met in just one hour.

Unfortunately, neither *E. coli* ArgE, *Sv*Dea or any other commercial acylases tested, were active towards the target *N*-Ac-PT, thus NAAAR DM activity towards this particular substrate could not be tested. This is in contrast with the results presented in a patent (US6177616, 2001). However, the NAAAR DM/ArgE couple was used in the DKR of *N*-Ac-NCAAs and the results show great potential for scaled up production of unnatural L-amino acids, such as propargylglycine (**2g**) and 1-triazolylalanine (**2i**). However, to ensure an efficient process (higher yields and lower reaction times), it is necessary to optimise the reaction conditions; so far, an incubation temperature of 50°C has given the highest conversions at the expense of ArgE stability in the reaction mixture, although overnight precipitation was observed. A possible solution to overcome the different thermostability of NAAAR and ArgE would consist in the immobilization of the two enzymes, either on the same or on separate matrixes.[196,197]

The substrate scope of NAAAR QM was also investigated, by setting up a series of small scale DKRs, but given the preference of this racemase for aromatic *N*-Ac-AAs, the acylases from Dr. Reddy's Laboratories were used in the screening. The NAAAR QM can be efficiently coupled with either a *T. litoralis* L-Acylase or *A. xylosoxidans* D-Acylase for the production of

enantiopure amino acids and is able to accept a range of *N*-Ac-aromatic or aliphatic amino acids, but shows low tolerance with polar and sterically hindered substrates, such as: *N*-Ac-Cit (**2t**) and *N*-Ac-BpyA. Furthermore, BpyA is known to be a good ligand for binding to transition metals, used for the preparation of artificial metallo-enzymes.[17] It is possible that *N*-Ac-BpyA, in a ten-fold excess compared to $CoCl_2$, coordinates with the cobalt present in the reaction mixture, essential for NAAAR activity, deactivating the racemase. Future work could involve repeating the reactions using a greater excess of $CoCl_2$ and lower *N*-Ac-BpyA concentration. To be certain that NAAAR is not able to accept this substrate, it would be necessary to monitor directly the racemization step starting from either *N*-Ac-L-BpyA, or *N*-Ac-D-BpyA. However, since no enantiopure samples were available, it can only be concluded with certainty that neither the *T. litoralis* L-Acylase nor the *A. xylosoxidans* D-Acylase from Dr. Reddy's Laboratories show activity for this substrate. In terms of enantioselectivity the NAAAR QM/ *T. litoralis* L-Acylase couple gave better results than the NAAAR QM/ *A. xylosoxidans* D-Acylase couple. It should be noted that this is independent of NAAAR, which simply catalyse the racemization equilibrium, but is related to the choice of the D-Acylase.

Given the positive results obtained from the initial screening, the NAAAR QM/Acylase couple was exploited to perform scale up DKR of four substrates. The average to high isolated yields of enantiopure amino acids produced, combined with the low NAAAR loading and the easy reaction set-up and product isolation protocol, make this process ideal for industrial application. However, the only limitation would be the poor enantioselectivity of the *A. xylosoxidans* D-Acylase used. This problem can be easily overcome either through testing various enzyme concentration or by choosing another acylase with better selectivity.

6 Materials and Methods

6.1 Materials and Reagents

6.1.1 Overview

Reagents, chemicals and media were purchased from Sigma-Aldrich, Thermo Fisher Scientific or Bio-Rad and used without further purification. Competent cells, pET plasmids and restriction enzymes were purchased from New England Biolabs (NEB), Novagen and Life Technologies. Primers were purchased from Sigma-Aldrich, while chromatography columns were obtained from GE Healthcare.

6.1.2 Competent cells

Various *E. coli* cell lines were used to store and amplify plasmids, during cloning and for the production of recombinant proteins (Table 6.1).

Cell Line	Supplier	Application
DH5 α	New England Biolabs (NEB)	Amplification and storage of plasmid DNA
C2987	Life Technologies	Plasmid storage and amplification during cloning and mutagenesis (high transformation efficiency)
BL21 (DE3)	New England Biolabs (NEB)	Protein expression host
Rosetta (DE3)	Novagen	Protein expression host

Table 6.1: *E. coli* strains used in this work.

6.1.3 Plasmids

Various plasmids (Table 6.2) were used during cloning to obtain recombinant proteins with various affinity tags. The pETHisTEV was gifted to us from Prof Jim Naismith, Oxford University.

Plasmid	Resistance	Application
pET28a	Kanamycin	Expression of proteins with a N-terminal 6xHis tag or no affinity tag
pETHisTEV	Kanamycin	Expression of proteins with a N-terminal Tobacco Etch Virus (TEV) 6xHis tag
pEBSRCTEVC10HIS	Ampicillin	Expression of proteins with a C-terminal 10xHis tag
pET20b	Ampicillin	Expression of proteins with no affinity tag
pET22b	Ampicillin	Expression of proteins with no affinity tag

pET26b	Kanamycin	Expression of proteins with no affinity tag

Table 6.2: List of expression plasmid used in this work.

6.1.4 Growth media, antibiotics and induction

Media was prepared by dissolving the components (Table 6.3) in distilled water and was then sterilized in the autoclave (20 min at 120°C). For the terrific broth the buffer solution was prepared and sterilized separately and combined with the media just before use.

Antibiotic stocks were prepared at a concentration of: 30 mg/mL for kanamycin, 100 mg/mL for ampicillin and 35 mg/mL for chloramphenicol. The ampicillin and kanamycin stocks were dissolved in distilled water (dH_2O), while chloramphenicol was dissolved in ethanol; all stocks were stored at -20°C.

During protein expression the desired antibiotic was added to the cooled sterilized media with a 1:1000 dilution from the stock solution for a final concentration of: 30 μg/mL for kanamycin, 100 μg/mL for ampicillin and 35 μg/mL for chloramphenicol.

Induction was carried out by diluting (1:1000 ratio) an isopropyl β-D-1-thiogalactopyranoside (IPTG) stock of the appropriate concentration (0.1-1.0 M) in the growth solution. All stocks were stored at -20°C.

Media	Ingredients
Luria Bertani (LB)	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl
LB Agar	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar
Super Optimal Broth (SOC)	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 2% w/v glucose
Sorbitol media	LB components were dissolved in 100 mM Na phosphate pH 7.5 along with 0.5 M sorbitol
Terrific Broth (TB)	12 g/L tryptone, 24 g/L yeast extract, 5 g/L glycerol (dissolve in 900 mL dH_2O), 23.1 mg/L KH_2PO_4 and 125.4 mg/L K_2HPO_4 (dissolve in 100 mL dH_2O).

Table 6.3: List of media used in this study and their recipes.

6.1.5 Buffers

Buffers were prepared by dissolving all components in distilled water and adjusting the pH to the desired value, using either hydrochloric acid (HCl) or sodium hydroxide (NaOH). Prior to chromatography all buffers were filtered through a 0.22 filter and degassed by sonication.

Phosphate buffer saline (PBS, pH = 7.4) is obtained by dissolving 5 tablets (Sigma) in 1 L of dH_2O, for a final concentration of KH_2PO_4 0.24 g/L, Na_2HPO_4 1.44 g/L, KCl 0.2 g/L and NaCl 8.0 g/L.

Buffer	Components
ArgE lysis	50 mM Tris pH 8.0, 150 mM NaCl, 20 mM Imidazole
ArgE elution	50 mM Tris pH 8.0, 150 mM NaCl, 500 mM Imidazole
ArgE storage	50 mM Tris pH 8.0, 150 mM NaCl
NAAAR lysis and storage	50 mM sodium phosphate pH 8.0, 100 mM NaCl
NAAAR elution	50 mM sodium phosphate pH 8.0, 1.0 M NaCl
TEV lysis	PBS, 300 mM NaCl, 10 mM imidazole
TEV elution	PBS, 300 mM NaCl, 500 mM imidazole
TEV storage	50 mM Tris pH 8.0, 300 mM NaCl, 50% glycerol
***Sv*Dea lysis**	0.1 M CAPS pH 9.5, 150 mM NaCl, 20 mM Imidazole
***Sv*Dea elution**	0.1 M CAPS pH 9.5, 150 mM NaCl, 500 mM Imidazole
***Sv*Dea storage**	0.1 M CAPS pH 9.5, 150 mM NaCl, 10% glycerol

Table 6.4: List of buffers and their components used for protein purification in this project.

6.2 Molecular Biology

The *Escherichia coli* ArgE gene (cloned into a pET28a plasmid) and *Streptomyces viridochromogenes* Tü494 Dea (cloned into a pUC57 plasmid) were purchased from GenScript.

6.2.1 Site Directed Mutagenesis

All primers (see Appendices) for site directed mutagenesis (SDM) were designed following the overlapping method.[170] All SDM PCR reactions (V_f = 50 µL) were set up as reported in Table 6.5 and the following thermal cycle was used to carry out the reaction: 98°C x 30 s, (98°C x 30 s, 65°/55°C x 30 s, 72°C x 3 min) x 30, 72°C x 10 min.

Reagents	Volumes (µL)
Phusion buffer 5x	10
dNTPs 10 mM	1.0
Fwd primer 10 µM	2.5
Rev primer 10 µM	2.5
DNA template (~100 ng)	1.0
DMSO	1.5
Phusion DNA polymerase (NEB)	1.0
dH_2O	up to 50

Table 6.5: Quantities of reagents used in an SDM reaction.

25 µL of reaction product were digested overnight at 37°C to remove the methylated parental DNA. The digestion reaction contained 1 µL DpnI (NEB), 3 µL 10xCustsmart buffer (NEB), 25 µL reaction mixture and 1 µL dH_2O; 5 µL of the digested mixture were transformed in *E.coli* C2987 cells and incubated overnight at 37°C (see Section 6.2.8).

Few colonies were selected for DNA isolation (see Section 6.2.9) and the presence of the desired mutation was confirmed from the sequencing results (see Section 6.2.10).

6.2.2 ArgE Directed Evolution

6.2.2.1 MEGAWHOP-PCR protocol

The following primers were used to set up an error prone PCR reaction,[192] using the conditions reported in Table 6.6 and the following thermal cycle: 98°C x 2 min, (98°C x 30 s, 50°/55°/60°/65°C x 30 s, 72°C x 1.5 min) x 25, 72°C x 10 min.

ArgE Fwd: ATGTCGTACCATCACCATCACCATCACGATTACGAT

ArgE Rev: CGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTTA

Reagents	Volumes (µL)
Mutagenic buffer 10x	5.0
EP dNTPs 10x	5.0
Fwd primer 25 µM	1.0
Rev primer 25 µM	1.0
DNA template (~100 ng)	1.0
$MnCl_2$ 12.5 mM	2.0
Taq DNA polymerase (NEB)	0.5
dH_2O	up to 50

Table 6.6: Quantities of reagents used in an error-prone reaction.

This PCR product was used as template to run a High-Fidelity PCR reaction (Table 6.7) to amplify the mutagenic product, using the following thermal cycle: 98°C x 2 min, (98°C x 30 s, 50°/55°/60°/65°C x 30 s, 72°C x 1.5 min) x 25, 72°C x 10 min.

Reagents	Volumes (µL)
Phusion buffer 5x	10
dNTPs 10 mM	1.0
Fwd primer 10 µM	2.5
Rev primer 10 µM	2.5
error-prone PCR product	1.0
Phusion DNA polymerase (NEB)	1.0
dH_2O	up to 50

Table 6.7: Quantities of reagents used in a High-Fidelity PCR reaction.

The PCR product was purified *via* agarose gel (see Section 6.2.5) and cloned into a pETHisTev plasmid *via* a MEGAWHOP-PCR,[180,192] Table 6.8. The following optimized thermal cycle was employed: 68°C x 5 min, 98°C x 2min, (98°C x 10 s, 68°C x 3 min) x 35, 72°C x 5 min. 40 µL reaction mixture was digested with 1 µL DpnI (NEB), 3 µL 10xCustsmart buffer (NEB), and 1 µL dH_2O to remove the methylated parental DNA. The digestion reaction was incubated at 37°C overnight and 5 µL of digested mixture was transformed in *E.coli* BL21 (DE3) cells and incubated overnight at 37°C (see Section 6.2.8). The resulting colonies were screened for activity as described in Section 6.3.2.

Reagents	Volumes (µL)
Phusion buffer 5x	10
dNTPs 10 mM	1.0
pETHISTEV-ArgE template (~100 ng)	0.5
Amplified error-prone PCR product (~20 ng)	20
DMSO	1.5
Phusion DNA polymerase (NEB)	1.0
dH_2O	up to 50

Table 6.8: Quantities of reagents used in a MEGAWHOP PCR reaction.

6.2.2.2 Gibson Assemby protocol

The following primers were designed to amplify the ArgE gene.

ArgE Fwd: ATGTCGTACTACCATCACCATCACCATC

ArgE Rev: GTGGTGGTGGTGGTGCTCGAGTT

These primers were employed to set up an error prone PCR reaction, using the condition reported in Table 6.9 and the following thermal cycle: 95°C x 2 min, (95°C x 30 s, 55°/60°/65°/68°C x 30 s, 72°C x 1.5 min) x 30, 72°C x 10 min.

Reagents	Volumes (µL)
Mutagenic buffer 10x	5.0
EP dNTPs 10x	5.0
Fwd primer 25 µM	1.0
Rev primer 25 µM	1.0
DNA template (~100 ng)	1.0
$MnCl_2$ 12.5 mM	2.0
Taq DNA polymerase (NEB)	0.5
dH_2O	up to 50

Table 6.9: Quantities of reagents used in an error-prone reaction.

While in parallel a High-Fidelity PCR (see Table 6.10) was set up to amplify the pETHisTEV plasmid backbone, using the following primers. Thermocycle conditions: 98°C x 2 min, (98°C x 30 s, 50°/55°/60°/65°C x 30 s, 72°C x 1.5 min) x 30, 72°C x 10 min.

pETHisTEV Fwd: TAACTCGAGCACCACCACCACCA

pETHisTEV Rev: GTGATGGTGATGGTAGTACGACA

Reagents	Volumes (µL)
Phusion buffer 5x	10
dNTPs 10 mM	1.0
Fwd primer 25 µM	1.0
Rev primer 25 µM	1.0
DNA template (~100 ng)	1.0
DMSO	1.5
Phusion DNA polymerase (NEB)	1.0
dH_2O	up to 50

Table 6.10: Quantities of reagents used in a high-fidelity reaction.

Both PCR products were purified by agarose gel extraction (see Section 6.2.5) and ligated together using the Gibson Assembly protocol (see Section 6.2.7.2), with a 1:2 molar ratio of vector to insert. 5 µL ligation reaction were used to transform an aliquot of *E. coli* C2987.

All colonies obtained were picked with a sterile loop and re-suspended in 600 µL of sterile water. DNA was extracted and used to transform *E.coli* BL21 (DE3) cells. The resulting colonies were used for selection.

6.2.3 *Sv*Dea Subcloning

The following primers were employed to introduce the NcoI restriction site at the N-terminus of the *Svdea* gene.

***Sv*Dea Fwd**: CCATCCATGGCGAGCAGCGAA

***Sv*Dea Rev**: TGGTGCTCGAGTTAGCGTTCGCC

A High-Fidelity PCR (Table 6.11) was run using this thermal cycle conditions: 98°C x 2 min, (98°C x 30 s, 55°/65°C x 30 s, 72°C x 1.0 min) x 30, 72°C x 10 min.

The presence of the PCR product was confirmed by running an analytical agarose gel using 5 µL of reaction mixture. The remaining 45 µL was cleaned using the ZYMO Research DNA clean & concentrator kit, following the manufacture protocol and digested overnight at 37°C at the

NcoI and XhoI restriction sites, along with 45 µL of empty pETHisTEV plasmid. Each reaction (50 µL) contained 45 µL of DNA, 2 µL of each restriction enzyme, and 1 µL of CutSmart buffer The digested fragments were purified *via* agarose gel and ligated (see Section 6.2.7.1). 5 µL of ligation reaction was transformed in C2987 cells.

Reagents	Volumes (µL)
Phusion buffer 5x	10
dNTPs 10 mM	1.0
Fwd primer 25 µM	1.0
Rev primer 25 µM	1.0
DNA template (~100 ng)	1.0
DMSO	1.5
Phusion DNA polymerase (NEB)	1.0
dH_2O	up to 50

Table 6.11: Quantities of reagents used in a high-fidelity reaction.

6.2.4 ArgE Subcloning

The pET28a-ArgE construct was digested, along the empty pETHisTEV plasmid at the NcoI and XhoI restriction sites overnight at 37°C. Each reaction (50 µL) contained 45 µL of DNA (~100 ng/µL), 2 µL of each restriction enzyme, and 1 µL of CutSmart buffer. The DNA fragments were purified *via* agarose gel and ligated; 5 µL of reaction mixture was transformed into C2987 cells.

6.2.5 Agarose gel electrophoresis and extraction

The gels were prepared mixing 1.5 g of agarose in 150 mL of TAE buffer (40 mM Tris, 20 mM acetic acid and 1 mM EDTA). The suspension was heated until the agarose completely dissolved and then cooled-down to 40-50°C, before adding GelRed (10 µL). The solution was poured in a casting mold and allowed to polymerize. The gel was loaded with the marker (Hyperladder I, Bioline) and the PCR products or other DNA samples, mixed with the 6x Purple DNA loading dye (NEB). The gel was run for 1 h at a constant voltage of 100 V. The DNA samples were visualized under UV-light and the bands of interest were cut from the gel using a scalpel and extracted using a Qiagen QIAquick gel extraction kit, following the manufacture protocol.

6.2.6 Digestion

DNA samples (40 µL) were mixed with CutSmart buffer (5 µL), two appropriate restriction enzymes (2 µL) and water (1 µL). The mixture was incubated at 37°C overnight and the product purified *via* Agarose gel electrophoresis.

6.2.7 Ligation

6.2.7.1 T4 DNA ligase protocol

A ratio of 1:3 of backbone (50 ng) to insert was selected to perform the ligation reaction (100 ng total DNA). The backbone and insert were mixed with 10x T4 DNA ligation buffer (1 µL) and T4 DNA ligase (1 µL), and the volume was brought up to 10 µL using water. The reaction was incubated at room temperature for 15 min and 5 µL of the reaction mixture was used to transform an aliquot of C2987 cells.

The following equation was used to calculate the amount of insert DNA needed in the ligation reaction with a 1:3 ratio:

$$\text{insert (ng)} = \frac{\text{vector (ng)} \times \text{size of insert (kbp)}}{\text{size of vector (kbp)}} \times \frac{3}{1}$$

6.2.7.2 Gibson Assembly protocol

A 2-fold molar excess of insert to backbone (100 ng) was used to perform a ligation by mixing together the insert and vector with the Gibson Assembly master mix (10 µL, NEB). Water was used to bring the reaction volume up to 20 µL. The reaction was incubated at 50°C for 15 min and 2 µL of reaction mixture was used to transform into C2987 cells.

The following equation was used to calculate the pmol of DNA needed for the ligation:

$$DNA\ (pmol) = \frac{\text{DNA (ng)} \times 1000}{\text{size of fragment (bp)} \times 650\ \text{Da}}$$

6.2.8 Heat shock transformation

Plasmid DNA (2-5 µL) was mixed with an aliquot of *E. coli* competent cells and incubated on ice for 30 minutes. The cells were heat shocked at 42°C for 45 min and immediately put back in ice for an additional 2 min. SOC (100 µL) was added to the cell and the mixture incubated at 37°C for 1 h on a shaker (200 rpm). After incubation the solution was spread on the surface of an LB-agar plate, containing the appropriate antibiotic. The plate was incubated at 37°C overnight.

6.2.9 DNA miniprep

10 mL of LB containing the appropriate antibiotic (10 µL) was inoculated with a single colony, picked from a fresh agar plate. The samples were incubated at 37°C for 18-20 h. The cells were harvested by centrifugation at 4000 rpm with a Heraeus Multifuge X3R equipped with a TX-1000 rotor for 15 min (4°C). The supernatant was discarded and the plasmid DNA extracted from cell pellets with a Qiagen QIAprep Miniprep kit, following the manufacture protocol. DNA concentration estimated using the nanodrop.

6.2.10 Analytical digest and sequencing

Incorporation of the gene of interest into the vector was checked after cloning by digesting the plasmid DNA (7 µL) with the appropriate restriction enzymes (1 µL each) and CutSmart buffer (1 µL). These samples were incubated at 37°C for 2 h and subsequently mixed with 6x Purple DNA loading dye (2 µL) and loaded onto an agarose gel. After the run the gel was checked using the UV-light; positive samples have two band: one at the size of the insert and one at the size of the vector. To confirm the gene sequence the positive samples were sequenced by GATC using the T7 and pET-RP primers.

6.3 Protein production

6.3.1 Expression test protocol

10 mL of LB with appropriate antibiotic was inoculated with a single colony of *E. coli* BL21 (DE3) cells. The culture was incubated for 18-20 h at 37°C on an orbital shaker (200 rpm). 250 µL of this overnight culture was inoculated into 10 mL of fresh media with appropriate antibiotic. These samples were incubated on an orbital shaker, until the optical density of the cells at 600 nm (OD_{600}) reached a value of 0.6-0.8 for LB media or 1.0-1.5 for TB media. The expression was induced by addition of IPTG at various concentrations and the cultures were incubated at 30°C for 4 h or 20°C for 18-20 h. Cultures were harvested by centrifugation at 4000 rpm, 4°C with a Heraeus Multifuge X3R equipped with a TX-1000 rotor for 15 min; the supernatant was discarded and the cell pellets re-suspended in 1mL of PBS buffer and lysed *via* sonication for 30 s (10 s ON/OFF). The soluble and insoluble fractions were separated by centrifugation with a benchtop centrifuge and the pellets re-suspended in 1 mL of PBS buffer. Level of expression of soluble and insoluble fractions was determined by analysis with a

denaturing 12% or 15% SDS-PAGE gel run at 200 V for 1 h. The optimal conditions determined were used for scale-up expression.

6.3.2 ArgE mutant library screening

In a sterile 24-well plate, colonies were picked into 2 mL LB media + kanamycin and incubated at 150 rpm at 37°C for 20 h. 100 µL of this overnight culture was inoculated into 2 mL of fresh media with kanamycin on a new 24-well plate, which was incubated at 37°C until OD_{600} reached a value of 0.6-0.8. Protein expression was induced by adding 0.2 mM IPTG and 0.1 mM $ZnSO_4$, the plate was incubated at 20°C overnight. Cultures were harvested by centrifugation at 4000 rpm, 4°C with a Heraeus Multifuge X3R equipped with a TX-1000 rotor for 40 min; the supernatant was discarded and the cell pellets re-suspended in 200 µL of NaPi 0.1 M pH 8 + 150 mM NaCl buffer and lysed by addition of 50 µL BugBuster primary amine free reagent. The plate was incubated at 25°C for 20 min and centrifuged at 4000 rpm, 4°C for 30 min. The cell free extract supernatant was employed to test the mutant library activity.

6.3.3 NAAAR variants expression[122]

250 mL of LB with antibiotic was inoculated with a single colony of *E. coli* BL21 (DE3) cells from a fresh plate, containing an expression plasmid with a variant of the *naaar* gene. The culture was incubated for 18-20 h at 37°C on an orbital shaker (200 rpm). The OD_{600} was measured and this culture was diluted into 1 L of TB with appropriate antibiotic to a final OD_{600} of 0.1. The new cultures were grown at 37°C until their OD reached a value of 1.0-1.5. At that point the protein expression was induced by addition 0.4 mM IPTG. The cultures were incubated at 20°C for 20 h before harvesting. Cultures were harvested by centrifugation at 4000 rpm, 4°C with a Heraeus Multifuge X3R equipped with a TX-1000 rotor for 15 min. The supernatant was discarded and the pellets stored at -20°C.

6.3.4 ArgE Expression

250 mL of LB with antibiotic was inoculated with a single colony of *E. coli* BL21 (DE3) cells from a fresh plate, containing an expression plasmid with a variant of the *arge* gene. The culture was incubated for 18-20 h at 37°C on an orbital shaker (200 rpm). The optical density of the cells at 600 nm (OD_{600}) was measured and this culture was diluted into 1 L of LB with appropriate antibiotic to a final OD_{600} of 0.1. The new cultures were grown at 37°C until their OD reached a value of 0.6-0.8. At that point the protein expression was induced by addition 0.2 mM IPTG and 0.1 mM $ZnSO_4$. The cultures were incubated at 20°C for 20 h before

harvesting. Cultures were harvested by centrifugation at 4000 rpm, 4°C with a Heraeus Multifuge X3R equipped with a TX-1000 rotor for 15 min. The supernatant was discarded and the pellets stored at -20°C.

6.3.5 *Sv*Dea Expression

250 mL of LB with antibiotic was inoculated with a single colony of *E. coli* BL21 (DE3) cells from a fresh plate, containing an expression plasmid with a variant of the *Svdea* gene. The culture was incubated for 18-20 h at 37°C on an orbital shaker (200 rpm). The optical density of the cells at 600 nm (OD_{600}) was measured and this culture was diluted into 1 L of TB with appropriate antibiotic to a final OD_{600} of 0.1. The new cultures were grown at 37°C until their OD reached a value of 1.0-1.5. At that point the protein expression was induced by addition 0.2 mM IPTG. The cultures were incubated at 20°C for 20 h before harvesting. Cultures were harvested by centrifugation at 4000 rpm, 4°C with a Heraeus Multifuge X3R equipped with a TX-1000 rotor for 15 min. The supernatant was discarded and the pellets stored at -20°C.

6.3.6 TEV Expression

250 mL of LB with antibiotic was inoculated with a single colony of *E. coli* Rosetta cells from a fresh plate, containing the pET28a-TEV construct. The culture was incubated for 18-20 h at 37°C on an orbital shaker (200 rpm). The optical density of the cells at 600 nm (OD_{600}) was measured and this culture was diluted into 1 L of LB with appropriate antibiotic to a final OD_{600} of 0.1. The new cultures were grown at 37°C until their OD reached a value of 0.6-0.8. At that point the protein expression was induced by addition 0.4 mM IPTG. The cultures were incubated at 20°C for 20 h before harvesting. Cultures were harvested by centrifugation at 4000 rpm, 4°C with a Heraeus Multifuge X3R equipped with a TX-1000 rotor for 15 min. The supernatant was discarded and the pellets stored at -20°C.

The purification of TEV is carried out in three steps: cell free extract preparation *via* sonication, followed by IMAC with a HisTrap FF 1 mL column and overnight dialyses in storage buffer. The TEV sample is concentrated to 2 mg/mL and stored at -80°C.

6.4 Protein purification

All purification steps were carried out at 4°C or on ice on an AKTA Purifier machine, unless otherwise stated. Samples (20 µL) from all steps were retained for SDS-PAGE analysis.

6.4.1 Cell free extract preparation

6.4.1.1 Sonication

Cells were resuspended in buffer and lysed by sonication for 15 cycles (30 sec on, 30 sec off). The lysed cell suspension was clarified by centrifugation (Thermo Scientific Multifuge X3R) at 10000 rpm for 45 min at 4 °C and the supernatant filtered on 0.45 µM syringe filters to obtain the cell free extract.

6.4.1.2 Chemo-lyses

Cells were resuspended in buffer and lysed by treatment with lysozyme (from chicken egg white, Sigma) 10 mg/ g of pellet. The suspension was gently stirred for ~20 h at 4°C, following incubation the lysate was centrifuged at 10000 rpm for 45 min at 4 °C and the supernatant recovered for use in further purification steps.

6.4.2 Anion exchange

The cell-free extract was loaded on a HiPrep Q FF 16/10 (GE Healthcare) column. The column was washed with 5 column volumes (CV) of lysis buffer, then the protein was eluted with the elution buffer using a step gradient of NaCl (0.1-1 M), 0-45% over 9 CV and 45-100% over 2 CV, at 2.0 mL/min.[121] Fractions containing the protein were mixed together and concentrated with a MW cut off (MWCO) concentrator (Sartorius) to a final volume of 1-5 mL.

6.4.3 Nickel NTA purification

6.4.3.1 HisTrap protocol

The cell-free extract was loaded onto a 1 mL HisTrap nickel affinity column (GE Healthcare). The column was washed with lysis buffer for 5 CVs, then the protein eluted with the elution buffer using a linear gradient of imidazole (20 to 500 mM) over 30 CVs. Fractions containing the protein were mixed together and concentrated with a MWCO concentrator (Sartorius) to a final volume of 1-5 mL.

6.4.3.2 Manual HisTrap column

Ni^{2+} resin was washed 3-5 times with the lyses buffer and added to the cell free extract. The suspension was stirred at 4°C on an orbital shaker (150 rpm) for 2 h to allow the His-tagged protein to bind to the resin. The suspension was packed on a column, washed with lysis buffer

and the protein eluted using the elution buffer (500 mM imidazole). Fractions collected during elution were tested with Bradford assay to identify which sample contained the enzyme. These fractions were mixed together, concentrated with a MWCO concentrator (Sartorius) to a final volume of 1-5 mL and filtered on 0.45 µM syringe filters.

6.4.4 TEV cleavage

The His-tagged protein solution was incubated with 2 mg/mL of pure TEV protease and dialysed against storage buffer 2L for 16 h at 4°C. Following dialysis, the protein was injected into a 1 mL HisTrap nickel affinity column, the untagged enzyme was eluted with 20 CV of storage buffer, then the TEV protease was washed off the column using elution buffer (500 mM imidazole). The collected non-tagged protein is concentrated with a MWCO concentrator (Sartorius) to a final volume of 1-5 mL

6.4.5 Size exclusion chromatography

The concentrated enzyme (1-5 mL) was loaded onto a pre-equilibrated (storage buffer) HiPrep™ 16/600 size exclusion column (120 mL). Recombinant protein was eluted at a flow rate of 0.5 or 1 mL min^{-1} (based on the column backpressure). The fractions containing clean protein of interest were combined and frozen at -80 °C (NAAAR requires a 20% glycerol buffer for storage).

Enzyme	SEC column
E. coli ArgE	16/60 Superdex™ S-200
S. viridochromogenes Dea	16/60 Superdex™ S-75
Amycolaptopsys NAAAR	16/60 Sephacryl™ S-300

Table 6.12 Pre-calibrated SEC columns employed for enzyme purification in this work.

6.5 Protein characterization

6.5.1 Mass spectrometry

Protein Liquid Chromatography-Mass Spectrometry (LC-MS/ESI) analyses were carried out on a SYNAPT G2 mass spectrometer. The source was an electrospray ionization (ESI). 10 µL protein (10-20 µM) was injected and eluted following a gradient 5-95% MeCN + 0.1% formic

acid in water + 0.1% formic acid over a 12 mins period. Mass spectra were analysed with MassLynx software.

6.5.2 SDS-PAGE electrophoresis

An average gel (16 mL) consisted of a 12% (6.9 mL H_2O, 4.8 mL of 40% acrylamide, 4 mL of 1.5 M Tris pH 8.8, 150 µL of 10% w/v SDS, 300 µL of 50 mg mL^{-1} ammonium peroxodisulfate and 20 µL TEMED) or 15% (5.6 mL H_2O, 6 mL of 40% acrylamide, 4 mL of 1.5 M Tris pH 8.8, 150 µL of 10% w/v SDS, 300 µL of 50 mg mL^{-1} ammonium peroxodisulfate and 20 µL TEMED) running gel and a 4% stacking gel (2.9 mL H_2O, 0.75 mL of 40% acrylamide, 1.25 mL of 0.5 M Tris pH 6.8, 50 µL of 10% w/v SDS, 100 µL of 50 mg mL^{-1} ammonium peroxodisulfate and 5 µL TEMED). Polyacrylamide gel electrophoresis (PAGE) samples were prepared by 1:1 dilution in 2x SDS loading buffer (2.5 mL Tris, 0.5 M pH 6.8, 2 mL glycerol, 4Ml of 10% w/v SDS, 1 mL β-mercaptoethanol and 0.5 mL of 0.1% w/v bromophenol blue). These samples were boiled at 100°C for 5-10 minutes and 10 µL loaded onto the gel and run for ~ 60 min at 200V/180mA. Protein was visualised by InstantBlue Coomassie Stain (Expedeon Ltd) and the mass estimated using LMW marker (GE Healthcare).

6.5.3 Protein concentration

Protein concentration of pure samples and cell free extracts were estimated with the nanodrop, measuring the absorbance at 280/260 nm. The extinction coefficient of the enzymes was calculated from the amino acid sequence using the ProtParam tool (https://web.expasy.org/protparam/) software.

Enzyme	ε ($M^{-1}cm^{-1}$)
E. coli N-His tag ArgE	43360
E. coli non-tagged ArgE	37400
S. viridochromogenes Dea	45170
Amycolaptopsys NAAAR	34170

Table 6.13 Extinction coefficients of the various constructs used to estimate protein concentration

6.5.4 ICP-MS

Pure protein at a concentration of 1 mg/mL was employed to estimate the zinc and cobalt content of the enzyme, while a sample of storage buffer was used as blank; the analysis was

carried out by Dr Lorna Eades (University of Edinburgh, School of Chemistry). Samples (1 mL) were analysed by ICP-MS using an Agilent 7500ce (with octopole reaction system), employing a RF forward power of 1540 W and reflected power of 1 W, with argon gas flows of 0.81 L min^{-1} and 0.20 L min^{-1} for carrier and makeup flows, respectively. Sample solutions were taken up into the Micro mist nebuliser by peristaltic pump at a rate of approximately 1.2 mL min^{-1}. Skimmer and sample cones were made of nickel.

The instrument was operated in spectrum multi-tune acquisition mode and three replicate runs *per* sample were employed. Isotopes ^{59}Co, ^{66}Zn and ^{68}Zn were analysed in 'no gas' mode (Argon only) and each mass was analysed in fully quantitative mode with three points *per* unit mass.

A range of calibration standards for each element were prepared using single element 1000 mg l^{-1} standards (Fisher Scientific UK LTD Bishop Meadow Road, Loughborough, Leicestershire LE11 5RG), diluted with 18.2 Ω deionised water. A 100-fold dilution of Merk ICP Multi element standard solution VI CertiPUR® was employed as a reference standard for each of the elements. An internal standard (100 ppb Rh) was spiked into each standard and sample to correct for any instrument drift or matrix changes. Samples were diluted prior to analysis using18.2 Ω deionised water.

Parameters for no gas mode:

Ion Lenses:	**Quadrupole Parameters:**
Extract1: 0 V	OctP Bias: -6 V
Extract 2: -133 V	QP Bias: -3 V
Omega Bias-ce: -20 V	
Omega Lens-ce: 0 V	
Cell Entrance: -30 V	
QP focus: 3 V	
Cell Exit: -34 V	

6.6 Assays protocols

6.6.1 The L-AAO assay

The high-throughput L-AAO assay[122] was employed to determine the activity and kinetic parameters of the deacetylase and NAAAR enzymes. The data was elaborated with OriginLab (all experiments were repeated in triplicates).

For ***E. coli* ArgE** the kinetic analysis was carried out with a reaction mixture containing various concentration of substrate, ArgE 2.0-0.2 µM, $CoCl_2$ 100 µM, HRP (Sigma Aldrich) 10 U/mL, LAAO from Crotalus atrox (Sigma Aldrich) 70-250 mU/mL and *o*-dianisidine 0.1 mg/mL in 100 mM NaPi buffer pH 8. The reactions were incubated at 40°C and the absorbance at 436 nm was measured at intervals of 30 s for 1 hour. The concentration of ArgE and L-AAO was adjusted considering the enzymes activity towards the substrate used.

For ***Amycolaptopsis* NAAAR** the kinetic analysis was carried out with a reaction mixture containing various concentration of substrate, NAAAR 0.50 µM, Dr Reddy's L-Acylase 1.25 mg/mL, $MgCl_2$ 5 mM, HRP (Sigma Aldrich) 10 U/mL, LAAO from Crotalus atrox (Sigma Aldrich) 70 mU/mL and *o*-dianisidine 0.1 mg/mL in 100 mM NaPi buffer pH 8. The reactions were incubated at 50°C and the absorbance at 436 nm was measured at intervals of 30 s for 1 hour.

6.6.2 The *p*-NO_2 phenol assay[185]

The colorimetric *p*-NP assay was employed to test *S. viridochromogenes* Dea activity and determine its kinetic parameters for various *p*-NP esters. The substrate stocks are prepared in AcCN and the final reaction mixture contains various concentration of substrate and *Sv*Dea 100-50 µg/mL in 100 mM NaPi buffer pH 8. The reactions were incubated at 50°C and the absorbance at 405 nm was measured at intervals of 30 s for 1 hour. The data was elaborated with OriginLab (all experiments were repeated in triplicates).

6.6.3 The OPA assay

The OPA assay protocol reported by Kang *et al.*[178] was employed.

The reactive mixture was formed by mixing 10 mL of OPA (Merck) 15 mM and *N*-Ac-L-Cys (Sigma) 30 mM in ethanol with 50 mL of NaBi 140 mM buffer pH 9.5. This solution can be conserved at 4°C in darkness for up to three days.

The derivatization proceeds by mixing in a 96-well plate 40 µL of an amino acid standard solution or reaction mixture with 40 µL of reactive mixture. The plate is slowly shacked for 30 s, then 100 µL of water is added and the fluorescence is recorded using a microplate reader (λ_{ex} = 340 nm and λ_{em} = 460 nm). Measures are repeated in triplicates.

6.6.4 Marfey's derivatization protocol[157]

50 µL of amino acid standard or reaction mixture were mixed with 100 µL of Marfey's reagent (1% w/v dissolved in Acetone) and 20 µL of 1 M $NaHCO_3$. The sample was heated at 60 °C for 30 min at 250 rpm in a Thermo Shaker (PCMT). After the sample cooled down to room temperature, 40 µL of 1 M HCl were added.

The samples were centrifuged for 10 minutes at 10000 rpm at r.t. and the supernatant (10 µL) was injected onto Luna 5u C18 RP-HPLC column (100 Å, 250 x 4.60 mm, Phenomenex). The LC gradient ran with a step gradient from 15% to 50 % AcCN in water with 0.1% TFA over 30 min and 50% to 15% of AcCN over 5 min, flow rate 1.0 mL min^{-1}, temperature: 25°C, detection at 340 nm. A standard sample is always run as reference.

6.7 Biotransformations

6.7.1 ArgE biotransformations protocol

500 µL reactions were set up using various *N*-acetyl-L-AAs (20 mM), *E. coli* ArgE 0.25, 0.50 or 1.00 mg/mL and $CoCl_2$ 100 µM in 100 mM NaPi pH 8. The reactions were incubated at 40°C, 250 rpm for 24 h. 20 µL $HCl_{conc.}$ were added to the reaction mixture to precipitate the enzyme, the samples were centrifuged for 10 min at 13000 rpm at r.t. and the supernatant analyzed by ^{1}H-NMR, LC-MS and HPLC.

6.7.2 *Sv*Dea hydrolysis of *N*-Ac-L-amino acids

500 µL reactions were set up using various *N*-acetyl-L-AAs (20 mM) and *S. viridochromogenes* *Sv*Dea 0.25, 0.50 or 1.00 mg/mL in 100 mM NaPi pH 8. The reactions were incubated at 50°C, 250 rpm for 24 h. 20 µL $HCl_{conc.}$ were added to the reaction mixture to precipitate the enzyme, the samples were centrifuged for 10 min at 13000 rpm at r.t. and the supernatant analyzed by ^{1}H-NMR and LC-MS.

6.7.3 NAAAR/deacetylase DKR

NAAAR DM and QM were overexpressed as reported in Chapter 6.3.3. Cell pellets were re-suspended in 50 mM TRIS-HCL pH 8.0 (around 5 gr in 25 mL of buffer, 20% w/v) and sonicated on ice during 10 minutes (10s ON, 10s OFF), following ultracentrifugation for 30 min at 10000 rpm. The clear supernatant was filtered through a 0.22 µM filter. The protein concentration of this cell-free extract (CFE) was quantified with the nanodrop. DKRs were set up in a V_f 500 µL with 50 mM racemic *N*-acetyl amino acids, 1 mg/mL NAAAR DM or QM CFE, L/D-acylase (2.5 mg/mL) and $CoCl_2$ (1 mM) in NaPi 0.1 M pH 8 and incubated at either 60° or 37°C, depending on the acylase used, for 24 h at 250 rpm. The reactions were quenched by addition of 10 µL of NaOH 5M, the enzyme was removed by centrifugation at 13000 rpm for 10 min at r.t. and the crude reaction mixtures analyzed by ^{1}H-NMR and HPLC.

6.7.4 Gram scale enzymatic reaction

NAAAR QM CFEs was prepared as described above. Reactions were carried out in conical flasks containing: 1 g *N*-acetyl-D,L-amino acids, 250 mg L/D-acylase, 5 mM $CoCl_2$ and 1 mg/mL NAAAR QM CFE in 25 mL 100 mM NaPi pH 8.0, V_f: 25 mL. After 24 h incubation at either 60° or 37°C, 250 rpm, depending on the acylase used, the reactions were quenched by addition of HCl_{conc} and the pH of the solution adjusted to 2. After removing the solvent under *vacuum*, the crude mixture was dissolved in the minimum amount of MeOH and the precipitated salts and enzymes were removed by filtration. The methanol phase was collected, concentrated under *vacuum* and the product precipitated by addition of an excess of acetone. Pure amino acid hydrochloride was isolated by filtration under *vacuum*, dried and analyzed *via* NMR, MS and HPLC.

6.8 Organic Synthesis

6.8.1 Deprotection of FMOC amino acids

FMOC-amino acid (3 g) was placed in a round bottom flask and dissolved in DMF (0.1 M). Following the addition of piperidine (10 eqv) the solution was left stirring at r.t. for 2 h. The reaction was quenched by removing the solvent *in vacuo* and the residue was dissolved in hexane with the formation of a white precipitate. This precipitate was collected and washed

first with hexane (20 mL x 3) and then with EtOH (20 mL x 3). The combined ethanolic phases were concentrated under *vacuum* to obtain pure amino acid as a white powder.

6.8.2 Ester hydrolysis

The amino acid ester derivative (500 mg) was placed in a round bottom flask and dissolved in HCl 6 M (0.1 M); this solution was stirred under reflux for 16 h. The reaction was quenched by removing the solvent under *vacuum* and the free amino acid was obtained as a white powder.

6.8.3 Amino acid acetylation

6.8.3.1 1st protocol (hydrophobic amino acids)[83]

In a round bottom flask the amino acid (2 g-200 mg) was dissolved in a 1:1 mixture of DMF and water (1.0 M). NEt_3 (1.2 eqv) and Ac_2O (2.2 eqv) were added dropwise to the solution under stirring. After 20 h at r.t. the solvent was removed under vacuum and the residues re-dissolved in water; the pH of the solution was adjusted to 10 with $NaCO_{3\ sat}$ and the mixture washed with Et_2O (20 mLx3). The organic phase discarded and the pH of the water phase was adjusted to 2 with HCl 2 M and the solution extracted with EtOAc (20 mLx3). The organic phase was dried under $MgSO_4$, filtered and the solvent removed under vacuum to give pure *N*-acetyl-amino acid.

6.8.3.2 2nd protocol (polar amino acids)[83]

In a round bottom flask the amino acid (2 g-200 mg) was dissolved in a 1:1 mixture of DMF and water (1.0 M). NEt_3 (1.2 eqv) and Ac_2O (2.2 eqv) were added dropwise to the solution under stirring. After 20 h at r.t. the solvent was removed under vacuum and the residues re-dissolved in water; the pH of the solution was adjusted to 10 with $NaCO_{3\ sat}$ and the mixture washed with Et_2O (20 mLx3). The organic phase discarded and the pH of the water phase was adjusted to 2 with HCl 2 M and the solvent removed under vacuum. The white solid was dissolved in the minimum amount of MeOH, the solution filtered and methanol removed *in vacuo* to give pure *N*-acetyl-amino acid.

6.8.3.3 3rd protocol (aromatic amino acid)[191]

In a round bottom flask the amino acid (1.5 g) and NaOH (2.0 eqv) were dissolved in a 1:3 mixture of acetone and water (0.1 M). Slowly Ac_2O (1.5 eqv) was added to the solution and after addition, the mixture was left stirring for 20 h at r.t. The reaction was quenched by

removal of the acetone under *vacuum*, $NaCO_{3\ sat}$ was added to the remaining solution until the pH reached a value of 10. The solution was then washed with Et_2O (20 mLx3) and the pH of the water phase was adjusted to 2 with HCl 2 M. This solution was extracted with EtOAc (20 mLx3) and the collected organic phases were dried under $MgSO_4$, filtered and the solvent removed under *vacuum* to give pure *N*-acetyl-amino acid.

6.8.4 Marfey's reagent synthesis[157]

L-alanine amide (0.131 g, 1.5 mmol) was dissolved into acetone (0.1 M); 1 mL of NaOH (1 M) was added to the mixture, which was left stirring for 5 min or until the starting material is completely dissolved. $MgSO_4$ (2.8 g) was added to the solution, which was stirred at r.t. for 3 h, before being filtered. The filtered solution was added dropwise into a second solution of di-nitro-di-fluoro benzene (173 mg, 0.85 mmol) in acetone (0.2 M). The reaction mixture was stirred at r.t. for 30 minutes and quenched by addition of cold water (20 mL) dropwise. The bright yellow needle shaped crystals formed are filtered under *vacuum*, washed with cold water, dried and stored in darkness.

6.9 Compound Characterization

6.9.1 NMR spectroscopy

Nuclear magnetic resonance (NMR) spectra were recorded at 298 K on Bruker PRO500 or AVA500 spectrometers running at 500 MHz, solvent peak taken as reference. Coupling constants J, were calculated using MestreNova. The following abbreviations (and their combinations) are used to label the multiplicities: s (singlet), d (doublet), t (triplet), q (quadruplet) and m (multiplet).

For reaction monitoring the NMR samples were prepared by adding 100 µL of D_2O to 400 µL of reaction mixtures and the NMR spectra recorded on a 500 MHz, single EspriSat experiment for water (16 scans, solvent peak suppression). Relative conversions measured by comparison of the integrals of the acetate peak of the substrate and the product.

$$Conv\ (\%) = \frac{(I_P)}{(I_P + I_R)} \cdot 100$$

6.9.2 LC-ESI MS spectrometry

The analysis was run on a Bruker micrOTOF LC-MS system with an ESI source. The samples for characterization were prepared by dissolving 5-10 mg materials in 1 mL MS grade water or methanol, while for reaction monitoring 50 µL of reaction mixture were diluted in 50 µL of H_2O and AcCN 1:1 mixture (MS grade solvents) and filtered before injection.

6.9.3 Chiral HPLC

6.9.3.1 Reaction monitoring

Time point samples were taken from an enzymatic reaction by diluting 20 µL of reaction mixture in 0.025% triethyl ammonium acetate (TEAA):MeOH (75:25) at different times. Samples were then analyzed by chiral HPLC using a Chirobiotic T column (Astec, chiral phase Teicoplanin, 5 µm, 250 mm x 4.6 mm) and the following isocratic conditions: mobile phase: 0.025% TEAA:MeOH (75:25, v/v), flow rate: 0.5 mL min^{-1}, detection λ: 210 nm, temperature: 40 °C, run time: 30 minutes.[121] The column was calibrated with single enantiomers of the free amino acids and their *N*-acetyl derivatives, reactions were run in triplicates. The Chromoleon software (Dionex) was used to measure the area under substrate and product peaks.

6.9.3.2 Enantiomeric excess calculation

Following reaction work-up, HPLC samples were prepared by 1:40 dilution of reaction mixture into the HPLC mobile phase, all samples were filtered before injection into the HPLC column (Chirobiotic T, Flow: 0.5 mL/min, 40°C, Mobile phase: 60:40 MeOH: 0.025% TEAA, protocol adapted from Turner *et.al.*[194]). Free amino acids standards and their *N*-acetyl derivatives were run for reference. The Chromoleon software (Dionex) was used to measure the area under substrate and product peaks. The enantiomeric excess (*ee*) was calculated using the area of the peaks for each enantiomer with the following formula:

$$ee\% = \frac{|(Area\ S - Area\ R)|}{(Area\ S + Area\ R)} \cdot 100$$

6.10 Crystal trials

This work was carried out in collaboration with Prof Gideon Grogan (University of York). Recombinant purified *E. coli* non-tagged ArgE WT and E144A mutant at 14 mg mL^{-1} was

crystallised by sitting-drop vapour diffusion at 18 °C. Crystallisation experiments were set up in 96 well MRC plates with commercially available screens from Molecular Dimensions Limited with 2/3 nL drops of protein and 2/3 nL well solution and equilibrated against 100 µL of well solution. Crystals grew in two conditions, CSS1+2 C12 (0.8 M Na formate, 0.1 M Tris pH 8, 8% PEG 20k and 8% PEG 550 MME) and D12 (0.8 M Na formate, 0.1 M Tris pH 8, 10% PEG 8k and 10% PEG 1k), which yielded prism shaped protein crystals with a 6 Å resolution. Various attempts were made to try to optimise these conditions, using both the sitting and hanging drop methods (4-5 µL drop size), however no diffracting quality crystals were obtained. More screening and optimisation are required. Datasets were collected on beamline I03 at the Diamond Light Source (Didcot, UK) at 100 K using a Pilatus 6M detector.

7 Conclusions

The aim of this thesis was to characterize two *N*-acetyl amino acids deacetylases with biocatalytic potential and couple them with an engineered *Amycolaptopsis* NAAAR. This would enable the development of an efficient biocatalytic dynamic kinetic resolution for the synthesis of L-phosphinothricin and other non-canonical amino acids.

E. coli ArgE is a co-catalytic zinc dependant metallo-deacetylase, involved in the biosynthesis of arginine in bacteria. This hydrolase was reported to be active towards *N*-Ac-PT, which makes the enzyme an ideal candidate for the DKR of this particular substrate. Following expression and purification of the acylase, ICP-MS and EDXRF analysis suggest the presence of only one Zn^{2+} ion bound for ArgE monomer and chiral HPLC confirmed the high stereoselectivity of the enzyme for the substrate L-isomer. The high-throughput colorimetric L-AAO assay was employed to optimize the reaction conditions, screen for activity and perform the kinetic analysis towards *N*-acetyl canonical and non-canonical amino acids.

A novel ^{1}H NMR assay was developed to monitor the deacetylation reaction and to calculate the conversions. ArgE did not show any activity toward *N*-Ac-PT, contrary to what was previously reported in literature. However, the wide substrate scope and high TONs and enantioselectivity suggest that the enzyme has a good biocatalytic potential for the synthesis of a wide range of NCAAs. Indeed, ArgE was efficiently coupled with NAAAR DM in a DKR of various aliphatic and heterocyclic amino acids. While the NAAAR QM, with enhanced activity for aromatic substrates, was coupled with two acylases with opposite stereoselectivity (obtained from Dr. Reddy's laboratories): *T. litoralis* L-Acylase and *A. xylosoxidans* D-Acylase, reported to be active towards these compounds.

A small-scale screening was run to study the QM substrate scope and eight DKRs were scaled up to produce enantiopure phenylalanine derivatives. The amino acids were isolated with an average yield ≥ 70%. Given the higher L-Acylase stereoselectivity, the four L-AAs produced were optically pure with an *ee* of 99.9%, while the D-AAs products suffered from low to average *ee*s (25-85%).

The crystal structure of ArgE has not been resolved. Following various rounds of optimization, the crystal trials performed had yielded protein crystals with a 4 Å resolution and the quality of the electron density map was not sufficient to resolve the enzyme structure. SDM studies

based on a homology model allowed to identify the residues directly involved in catalysis (E144) and intermediate stabilization (H195). With this, a random mutagenesis approach was employed to engineer ArgE and increase its activity although no positive hits have been isolated thus far.

S. viridochromogenes Dea is a novel deacetylase which catalyses the last step in the production of the herbicidal tripeptide bialaphos. In this work, the expression, purification and characterization of *Sv*Dea was reported for the first time. The ^{1}H NMR and two colorimetric assays, *p*NP and L-AAO, were employed to probe the enzyme substrate scope. The results obtained suggest that *Sv*Dea is selective towards its natural substrate yet the *N*-Ac-L-PTT hydrolysis requires to be scaled-up and repeated to obtain more conclusive data. The homology model reveals the use of a hydrolytic mechanism based on a catalytic triad, formed by S143, E237 and H267, while sequence analogy confirms that *Sv*Dea belongs to the HSL family suggesting that the enzyme may be involved in the catabolism of triglycerides.

This work mainly focused on the optimization of a biocatalytic process for the production of enantiopure amino acids. While a strereoseletive approach from prochiral precursors is the preffered strategy currently employed; the DKR here reported proved to be a good alternative given the high yields and *ee*s achieved. Furthermore, *Amycolatopis* NAAAR is a robust enzyme (high thermostability) with a broad substrate range, able to work cooperatly with various L- and D-selective aminoacylases, opening the road for the production of numerous usefull and potentially bioactive target compounds. With the increased interest in "safe" nature-derived agrochemicals, being able to prepare L-PT with a novel and environmentally friendly strategy would be advantageus. However, *N*-Ac-L-PT is not commonly accepted by many hydrolases as substrate, a broad screen of a large Acylase panel, followed by potentially engineering of both the NAAAR and the stereoselective hydrolase, may be required to obtain an efficient biocatalytic DKR applicable at industrial scale. Moreover, studying and understanding the mechanism of the enzymes involved in the biosynthesis of phosphinothricin could led to the development of novel and unexpected strategies for the production of this natural product.

Future work will mainly focus on using the error-prone PCR protocol optimised to evolve *E. coli* ArgE and widening the *S. viridochromogenes* Dea substrate scope by testing the enzyme activity for various acetylated esters or derivatives and scaling up the *N*-acetyl bialaphos hydrolysis. Further crystal trials will be set up using new mutants and/or constructs to obtain

good quality *Sv*Dea and ArgE crystals and to resolve the two deacetylase crystal structures. More studies will be conducted to study *Sv*Dea activity on fatty acids esters, and its involvement in the microorganism fatty acids metabolism. If an engineered version of ArgE active towards *N*-Ac-L-PT is isolated, this mutant will be coupled with the NAAAR variants to test the racemase activity towards this natural product and, if required, further mutagenesis will be employed to evolve the NAAAR enzyme.

8 References

1 N. Tonouchi and H. Ito, in *Amino Acid Fermentation*, eds. A. Yokota and M. Ikeda, Springer Japan, Tokyo, 2017, pp. 3–14.

2 F. Agostini, J. Vçller, B. Koksch, C. G. Acevedo-rocha, V. Kubyshkin and N. Budisa, *Angew. Chem. Int. Ed.*, 2017, **56**, 9680–9703.

3 H. Zou, L. Li, T. Zhang, M. Shi, N. Zhang, J. Huang and M. Xian, *Biotechnol. Adv.*, 2018, **36**, 1917–1927.

4 P. B. Nunn, Q. Mary and R. J. Nash, *NPC Nat. Prod. Commun.*, 2010, **5**, 485–504.

5 K. J. Rodgers, K. Samardzic and B. J. Main, in *(eds) Plant Toxins. Toxinology. Springer*, 2015, pp. 263–285.

6 M. A. T. Blaskovich, *J. Med. Chem.*, 2016, **59**, 10807–10836.

7 F. Parmeggiani, S. T. Ahmed, M. P. Thompson, N. J. Weise, J. L. Galman, D. Gahloth, M. S. Dunstan, D. Leys and N. J. Turner, *Adv.Synth. Catal.*, 2016, **358**, 3298–3306.

8 H. Zhou, M. Lijun, Y. Xinjian, L. Yayun, X. Gang, W. Jianping, W. Mianbin and Y. Lirong, *Eur. J. Org. Chem.*, 2019, **2019**, 6470–6477.

9 P. J. Almhjell, C. E. Boville and F. H. Arnold, *Chem. Soc. Rev.*, 2018, **47**, 8971–9160.

10 N. A. Mcgrath, M. Brichacek, J. T. Njardarson and U. States, *J. Chem. Educ.*, 2010, **87**, 1348–1349.

11 J. Zhao, A. J. Burke and A. P. Green, *Curr. Opin. Chem. Biol.*, 2020, **55**, 136–144.

12 C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413–444.

13 Y. Huang and T. Liu, *Synth. Syst. Biotechnol.*, 2018, **3**, 150–158.

14 Y. Ravikumar, S. P. Nadarajan, T. H. Yoo, C. Lee and H. Yun, *Biotechnol. J.*, 2015, **10**, 1862–1876.

15 A. J. Burke, S. L. Lovelock, A. Frese, R. Crawshaw, M. Ortmayer, M. Dunstan, C. Levy and A. P. Green, *Nature*, 2019, **570**, 219.

16 H. Xiao, F. B. Peters, P. Yang, S. Reed, J. R. Chittuluru and P. G. Schultz, *ACS Chem. Biol.*, 2014, **9**, 1092–1096.

17 I. Drienovská, A. Rioz-Martínez, A. Draksharapua and G. Roelfes, *Chem. Sci.*, 2015, **6**, 770–776.

18 T. Hayashi, D. Hilvert and A. P. Green, *Chem. Eur. J.*, 2018, **24**, 11821–11830.

19 M. D'Este, M. Alvarado-Morales and I. Angelidaki, *Biotechnol. Adv.*, 2018, **36**, 14–25.

20 N. Dilini, C. Theisen, T. Walter and K. D. Walker, *J. Biotechnol.*, 2016, **217**, 12–21.

21 H. Gröger, *Chem. Rev.*, 2003, **103**, 2795–2827.

22 C. Nájera and J. M. Sansano, *Chem. Rev.*, 2007, **107**, 4584–4671.

23 W. W. Metcalf and W. A. van der Donk, *Annu. Rev. Biochem.*, 2009, **78**, 65–94.

24 R. Finking and M. A. Marahiel, *Annu. Rev. Microbiol.*, 2004, **58**, 453–488.

25 N. Grammel, D. Schwartz, W. Wohlleben and U. Keller, *Biochemistry*, 1998, **37**, 1596–1603.

26 D. Schwartz, S. Berger, E. Heinzelmann, K. Muschko, K. Welzel and W. Wohlleben, *Appl. Environ. Microbiol.*, 2004, **70**, 7093–7102.

27 M. Alberto, M. Núñez and V. Eric, *Sustain. Chem. Process.*, 2016, **4**, 1–8.

28 I. Ujváry, in *Hayes' Handbook of Pesticide Toxicology*, Elsevier Inc., Third Edit., 2010, pp. 119–229.

29 X. Chen, W. Yang, E. Sivamani, A. H. Bruneau, B. Wang and R. Qu, *Mol. Breed.*, 2005, **15**, 339–347.

30 H. Unno, T. Uchida, H. Sugawara, G. Kurisu, T. Sugiyama, T. Yamaya, H. Sakakibara, T. Hase and M. Kusunoki, *J. Biol. Chem.*, 2006, **281**, 29287–29296.

31 W. Wohllebent and Y. Tatenovii, *Proc. Natl. Acad. Sci.*, 1993, **90**, 3009–3013.

32 K. Bartsch and C. C. Tebbe, *Appl. Environ. Microbiol.*, 1989, **55**, 711–716.

33 E. Strauch, W. Wohllehen and A. Puhler, *Gene*, 1988, **63**, 65–74.

34 B. Christ, R. Hochstrasser, L. Guyer, R. Francisco, S. Aubry, S. Hörtensteiner and J. Weng, *Nat. Plants*, 2017, **3**, 937–945.

35 J. M. Green and M. D. K. Owen, *J. Agric. Food Chem.*, 2011, **59**, 5819–5829.

36 H. K. Takano, R. Fernando, L. Ovejero, G. G. Belchior, G. Potrich and L. Maymone, *Sci. Agric.*, 2021, **78**, 1–11.

37 S. Duke, *Environ. Health Perspect.*, 1990, **87**, 263–271.

38 H. J. Zeiss, *J. Org. Chem.*, 1991, **56**, 1783–1788.

39 H. J. Zeiss, *Tetrahedron*, 1992, **48**, 8263–8270.

40 Lothar Willms, *US Pat.*, 2002, 6,359,162 B1.

41 M. Ruhland, G. Engelhardt and K. Pawlizki, *Pest Manag. Sci.*, 2004, **60**, 691–696.

42 U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore and K. Robins, *Nature*, 2012, **485**, 185–194.

43 U. T. Bornscheuer, *Philos. Trans. Ser. A, Math. Phys. Eng. Sci.*, 2018, **376**.

44 M. T. Reetz, *J. Am. Chem. Soc.*, 2013, **135**, 12480–12496.

45 P. Anastas and N. Eghbali, *Chem.Soc.Rev.*, 2010, **39**, 301–312.

46 P. A. Patten, R. J. Howard and W. P. C. Stemmer, *Curr. Opin. Biotechnol.*, 1997, **8**, 724–733.

47 A. Crameri, S.-A. Raillard, E. Bermudez and W. P. C. Stemmer, *Nature*, 1998, **391**, 288–291.

48 M. T. Reetz, *Directed Evolution of Selective Enzymes*, 2017.

49 F. Cadet, N. Fontaine, G. Li, J. Sanchis, M. Ng Fuk Chong, R. Pandjaitan, I. Vetrivel, B. Offmann and M. T. Reetz, *Sci. Rep.*, 2018, **8**, 16757.

50 K. Chen and F. H. Arnold, *Nat. Catal.*, 2020, **3**, 203–213.

51 S. Bähr, S. Brinkmann-Chen, M. Garcia-Borràs, J. M. Roberts, D. E. Katsoulis, K. N. Houk and F. H. Arnold, *Angew. Chemie Int. Ed.*, 2020, **59**, 15507–15511.

52 A. S. Bommarius, *Annu. Rev. Chem. Biomol. Eng.*, 2015, **6**, 319–345.

53 J. M. Choi, S. S. Han and H. S. Kim, *Biotechnol. Adv.*, 2015, **33**, 1443–1454.

54 E. M. M. Abdelraheem, H. Busch, U. Hanefeld and F. Tonin, *React. Chem. Eng.*, 2019, **4**, 1878–1894.

55 D. P. de Maria, G. de Gonzalo and A. R. Alcantara, *Catalysis*, 2019, **9**, 802.

56 D. L. Hughes, *Org. Process Res. Dev.*, 2018, **22**, 1063–1080.

57 J. P. Adams, M. J. B. Brown, A. Diaz-rodriguez, R. C. Lloyd and G. Roiban, *Adv.Synth. Catal.*, 2019, **361**, 2421–2432.

58 P. N. Devine, R. M. Howard, R. Kumar, M. P. Thompson, M. D. Truppo and N. J. Turner, *Nat. Rev. Chem.*, 2018, **2**, 409–421.

59 C. K. Saville, J. M. Janey, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz and J. Brands, *Science (80).*, 2010, **329**, 305–310.

60 R. O. B. Wilson, *Chem. Today*, 2015, **33**, 50.

61 M. A. Huffman, A. Fryszkowska, O. Alvizo, M. Borra-garske, K. R. Campos, K. A. Canada, P. N. Devine, D. Duan, J. H. Forstater, S. T. Grosser, H. M. Halsey, G. J. Hughes, J. Jo, L. A. Joyce, J. N. Kolev, J. Liang, K. M. Maloney, B. F. Mann, N. M. Marshall, M. Mclaughlin, J. C. Moore, G. S. Murphy, C. C. Nawrat, J. Nazor, S. Novick, N. R. Patel and A. Rodriguez-granillo, *Science (80).*, 2019, **366**, 1255–1259.

62 Y.-P. Xue, C.-H. Cao and Y.-G. Zheng, *Chem. Soc. Rev.*, 2018, **47**, 1516–1561.

63 K. Vedha-peters, M. Gunawardana, J. D. Rozzell and S. J. Novick, *J. Am. Chem. Soc.*, 2006, **128**, 10923–10929.

64 X. Yin, J. Wu and L. Yang, *Appl. Microbiol. Biotechnol.*, 2018, **102**, 4425–4433.

65 A. Gomm, S. Grigoriou, C. Peel, J. Ryan, N. Mujtaba, T. Clarke, E. Kulcinskaja and E. O'Reilly, *European J. Org. Chem.*, 2018, **2018**, 5282–5284.

66 I. Rowles, B. Groenendaal, B. Binay, K. J. Malone, S. C. Willies and N. J. Turner, *Tetrahedron*, 2016, **72**, 7343–7347.

67 B. Wu, W. Szymański, G. G. Wybenga, M. M. Heberling, S. Bartsch, S. de Wildeman, G. J. Poelarends, B. L. Feringa, B. W. Dijkstra and D. B. Janssen, *Angew. Chemie Int. Ed.*, 2012, **51**, 482–486.

68 C. L. Windle, M. Müller, A. Nelson and A. Berry, *Curr. Opin. Chem. Biol.*, 2014, **19**, 25–33.

69 M. Rachwalski, N. Vermue and F. P. J. T. Rutjes, *Chem.Soc.Rev.*, 2013, **42**, 9268–9282.

70 K. Faber, *Biotransformations in Organic Chemistry*, 2011.

71 O. Verho and J. Ba, *J. Am. Chem. Soc.*, 2015, **137**, 3996–4009.

72 S. A. Brown, M. Parker and N. J. Turner, *Tetrahedron: Asymmetry*, 2000, **11**, 1687–1690.

73 J. Qiu, E. Su, W. Wang and D. Wei, *Tetrahedron Lett.*, 2014, **55**, 1448–1451.

74 C. Femmer, M. Bechtold, T. M. Roberts and S. Panke, *Appl. Microbiol. Biotechnol.*, 2016, **100**, 7423–7436.

75 M. He, *J. Ind. Microbiol. Biotechnol.*, 2006, **33**, 401–407.

76 O. May, S. Verseck, A. Bommarius, K. Drauz, D. Ag, R. Chaussee, C. Engineering and F. D. V, *Org. Process Res. Dev.*, 2002, **6**, 253–257.

77 J. Altenbuchner, M. Siemann-herzberg and C. Syldatk, *Chem. Biotechnol. Fig.*, 2001, **12**, 559–563.

78 M. J. Rodríguez-Alonso, J. M. Clemente-Jiménez, F. Rodríguez-Vico and F. J. Las Heras- Vázquez, *Biochem. Eng. J.*, 2015, **101**, 68–76.

79 S.-H. Nam, H.-S. Park and H.-S. Kim, *Chem. Rec.*, 2005, **5**, 298–307.

80 O. May, S. Verseck, A. Bommarius, K. Drauz, D. Ag and R. Chaussee, *Org. Process Res. Dev.*, 2002, **6**, 253–257.

81 K. Yasukawa and Y. Asano, *Adv. Synth. Catal.*, 2012, **354**, 3327–3332.

82 H. K. Chenault, J. Dahmer and G. M. Whitesides, *J. Am. Chem. Soc.*, 1989, **111**, 6354.

83 B. Wang, Y. Liu, D. Zhang, Y. Feng and J. Li, *Tetrahedron: Asymmetry*, 2012, **23**, 1338–1342.

84 D. Shi, X. Yu, L. Roth, M. Tuchman and N. M. Allewell, *Biophys. Chem.*, 2007, **126**, 86–93.

85 S. Kumagai, M. Kobayashi, S. Yamaguchi, T. Kanaya, R. Motohashi and K. Isobec, *J. Mol. Catal. B*, 2004, **30**, 159–165.

86 J. M. Woodley, *Appl. Microbiol. Biotechnol.*, 2019, **103**, 4733–4739.

87 S. Tokuyama, K. Hatano and T. Takahashi, *Biosci. Biotech. Biochem.*, 1994, **58**, 24–27.

88 S. Tokuyama and K. Hatano, *Appl Microbiol Biotechnol*, 1995, **42**, 853–859.

89 S. C. Su and C. Lee, *Enzyme Microb. Technol.*, 2002, **30**, 647–655.

90 S. Verseck, A. Bommarius and R. Kula, *Appl Microbiol Biotechnol*, 2001, **55**, 354–361.

91 W. C. Wang, W. C. Chiu, S. K. Hsu, C. L. Wu, C. Y. Chen, J. S. Liu and W. H. Hsu, *J. Mol. Biol.*, 2004, **342**, 155–169.

92 M. Hayashida, S. Kim, K. Takeda, T. Hisano and K. Miki, *Proteins*, 2008, **1**, 519–523.

93 A. I. Martı, S. Martı, F. Rodrı, F. J. Las Heras-va and J. Marı, *Process Biochem.*, 2009, **44**, 835–841.

94 J. M. Clemente-jime, F. J. Las Heras-va and F. Rodrı, *Biopolymers*, 2009, **91**, 757–772.

95 J. Marı and F. J. Las Heras-va, *Mol Biotechnol*, 2015, **57**, 454–465.

96 S. Martínez-rodríguez, P. Soriano-maldonado and J. Antonio, *BBA - Proteins Proteomics*, 2020, **1868**, 140377.

97 D. R. J. Palmer, J. B. Garrett, V. Sharma, R. Meganathan, P. C. Babbitt and J. A. Gerlt, *Biochemistry*, 1999, **38**, 4252–4258.

98 J. A. Gerlt, P. C. Babbitt, M. P. Jacobson and S. C. Almo, *J. Biol. Chem.*, 2012, **287**, 29–34.

99 P. C. Babbitt, M. S. Hasson, J. E. Wedekind, D. R. J. Palmer, W. C. Barrett, G. H. Reed, I. Rayment, D. Ringe, G. L. Kenyon and J. A. Gerlt, *Biochemistry*, 1996, **36**, 16489–16501.

100 M. E. Glasner, N. Fayazmanesh, R. A. Chiang, A. Sakai, M. P. Jacobson, J. A. Gerlt and P. C. Babbitt, *J. Mol. Biol.*, 2006, **360**, 228–250.

101 D. M. Z. Schmidt, E. C. Mundorff, M. Dojka, E. Bermudez, J. E. Ness, S. Govindarajan, P. C. Babbitt, J. Minshull and J. A. Gerlt, *Biochemistry*, 2003, **42**, 8387–8393.

102 J. E. Vick, D. M. Z. Schmidt and J. A. Gerlt, *Biochemistry*, 2005, **44**, 11722–11729.

103 D. Odokonyero, A. W. Mcmillan, U. A. Ramagopal, R. Toro, D. P. Truong, M. Zhu, M. S. Lopez, B. Somiari, M. Herman, A. Aziz, B. Bonanno, K. G. Hull, S. K. Burley, D. Romo, S. C. Almo and M. E. Glasner, *Biochemistry*, 2018, **57**, 3676–3689.

104 M. Boersch and G. Grant, *RSC Adv.*, 2018, **8**, 5099–5105.

105 A. Sakai, D. F. Xiang, C. Xu, L. Song, W. S. Yew, F. M. Raushel and J. A. Gerlt, *Biochemistry*, 2006, **45**, 4455–4462.

106 A. Aliashkevich, L. Alvarez, F. Cava and L. Moe, *Front. Microbiol.*, 2018, **9**, 1–11.

107 J. B. Thoden, E. A. T. Ringia, J. B. Garrett, J. A. Gerlt, H. M. Holden and I. Rayment, *Biochemistry*, 2004, **43**, 5716–5727.

108 A. W. Mcmillan, M. S. Lopez, M. Zhu, B. C. Morse, I. Yeo, J. Amos, K. Hull, D. Romo and M. E. Glasner, *Biochemistry*, 2014, **53**, 4434–4444.

109 S. Tokuyama and K. Hatano, *Appl. Microbiol Biotechnol*, 1996, **44**, 774–777.

110 S. Hsu, H. Lo, C. Kao, D. Lee and W. Hsu, *Biotechnol. Prog.,* 2006, **22**, 1578–1584.

111 A. D. Bommarius, K. P. Drauz, M.-R. Kula and S. D. Verseck, *EP Pat.*, 2004, P1074628 B1.

112 S. Baxter, D. Campopiano and K. E. Holt-Tiffin, *US Pat.*, 2014, 20140065679 A1.

113 A. Bommarius, K. Drauz, S. Verseck and M.-R. Kula, *US Pat.*, 2004, 6,767,725 B2.

114 A. Bommarius, S. Verseck, K. Drauz and M.-R. Kula, *EP Pat. Pat.*, 2002, 10050124 A1.

115 H. Ikeda, Y. Yonetani, S. Hashimoto, M. Yagasaki and K. Soda, *US Pat.*, 2007, 7247466 B2.

116 A. Bommarius, K. Drauz and S. Verseck, *US Pat.*, 2003, US6656710 B2.

117 T. Takahashi and K. Hatano, *US Pat.*, 1991, 4981799.

118 A. Bommarius, K. Drauz and S. Verseck, *US Pat.*, 2007, 7157251 B2.

119 S. Verseck, M.-R. Kula, A. Bommarius and K. Drauz, *US Pat.*, 2002, 6372459 B1.

120 A. Bommarius, K. Drauz and S. Verseck, *US Pat.*, 2008, 7378269 B2.

121 S. Baxter, S. Royer, G. Grogan, F. Brown, K. E. Holt-Tiffin, I. N. Taylor, I. G. Fotheringham and D. J. Campopiano, *J. Am. Chem. Soc.*, 2012, **134**, 19310–19313.

122 G. Sánchez-Carrón, T. Fleming, K. E. Holt-Tiffin and D. J. Campopiano, *Anal. Chem.*, 2015, **87**, 3923–3928.

123 Y. Sumida, S. Iwai, Y. Nishiya and S. Kumagai, *Adv. Synth. Catal.*, 2016, **358**, 2041–2046.

124 Y. Sumida, S. Iwai, Y. Nishiya, S. Kumagai, T. Yamada and M. Azuma, *J. Biosci. Bioeng.*, 2018, **125**, 282–286.

125 S. Martínez-rodríguez, A. I. Martínez-gómez, F. Rodríguez-vico, J. M. Clemente-jiménez and F. J. Las Heras-vázquez, *Appl Microbiol Biotechnol*, 2010, **85**, 441–458.

126 S. Hsu, H. Lo, W. Lin, I. Chen, C. Kao and W. Hsu, *Process Biochem*, 2007, **42**, 856–862.

127 M. Yen, W. Hsu and S. Lin, *Process Biochem.*, 2010, **45**, 667–674.

128 P. Soriano-maldonado, F. J. Las Heras-vazquez, J. M. Clemente-jimenez, F. Rodriguez-vico and S. Martínez-rodríguez, *Appl. Microbiol. Biotechnol.*, 2015, **99**, 283–291.

129 P. Soriano-maldonado, M. J. Rodríguez-alonso, C. Hernández-cervantes, I. Rodríguez-garcía, J. M. Clemente-jiménez, F. Rodríguez-vico, S. Martínez-rodríguez, F. Javier and L. Heras-vázquez, *Process Biochem.*, 2014, **49**, 1281–1287.

130 U. Engel, J. Rudat and C. Syldatk, in *Industrial biocatalysis.*, Pan Stanford Publ., Singapore, 2014, vol. 1, pp. 817–862.

131 M. J. Rodríguez-Alonso, F. Rodríguez-Vico, F. J. Las Heras-Vázquez and J. M. Clemente-Jiménez, *J. Chem. Technol. Biotechnol.*, 2016, **91**, 1972–1981.

132 M. J. Rodríguez-Alonso, F. Rodríguez-Vico, F. J. Las Heras-Vázquez and J. M. Clemente-Jiménez, *Catalysis*, 2017, **7**, 192.

133 J. Enoki, J. Meisborn, A. Müller and R. Kourist, *Front. Microbiol*, 2016, **7**, 1–8.

134 H. S. Toogood, E. J. Hollingsworth, R. C. Brown, I. N. Taylor, S. J. C. Taylor, R. Mccague and J. A. Littlechild, *Extremophiles*, 2002, **6**, 111–122.

135 B. M. Parker, I. N. Taylor, J. M. Woodley, J. M. Ward and P. A. Dalby, *J. Biotechnol.*, 2011, **155**, 396–405.

136 J. Littlechild, *Front. Bioeng. Biotechnol.*, 2015, **3**, 161.

137 K. Bartsch, G. Kriete, I. Broer and A. Puhler, *US Pat.*, 2001, 6,177,616 B1.

138 A. P. and I. B. G. Kriete, K. Niehaus, A.M. Perlick, *Plant J.*, 1996, **9**, 809–818.

139 G. S. Rao, A. K. Tyagi and K. V. Rao, *Plant Sci.*, 2017, **256**, 139–147.

140 T. Meinnel, E. Schmitff and Y. Mechulam, 1992, **174**, 2323–2331.

141 F. Javid-Majd and J. S. Blanchard, *Biochemistry*, 2000, **39**, 1285–1293.

142 R. C. H. Ã, K. P. Bzymek and S. I. Swierczek, *Curr. Opin. Chem. Biol.*, 2003, **7**, 197–206.

143 R. H. Davis, 1986, **50**, 280–313.

144 S. Bonissone, N. Gupta, M. Romine, R. A. Bradshaw and P. A. Pevzner, *Mol. Cell. Proteomics*, 2013, **12**, 14–28.

145 W. C. McGregor, S. I. Swierczek, B. Bennett and R. C. Holz, *J. Am. Chem. Soc.*, 2005, **127**, 14100–14107.

146 S. Fruncillo, M. Trande, C. F. Blanford, A. Astegno and L. S. Wong, *Anal. Chem.*, 2019, **91**, 11502–11506.

147 J. Hirose, S. Ando and Y. Kidani, *Biochemistry*, 1987, **26**, 6561–6565.

148 F. Rigoldi, S. Donini, A. Redaelli and E. Parisini, *Appl. Bioeng.*, 2018, **2**, 1–17.

149 K. Huynh and C. L. Partch, *Curr. Protoc. Protein Sci.*, 2016, **28**, 1–19.

150 M. L. Dart, T. Machleidt, E. Jost, M. K. Schwinn, M. B. Robers, C. Shi, T. A. Kirkland, M. P. Killoran, J. M. Wilkinson, J. R. Hartnett, K. Zimmerman and K. V Wood, *ACS Med. Chem. Lett.*, 2018, **9**, 546–55.

151 F. Dupeux and M. Ro, *Acta Crystallogr.*, 2011, **D67**, 915–919.

152 J. M. Risse, A. Pühler and E. Flaschel, *Eng. Life Sci.*, 2005, **5**, 38–45.

153 S. Yokoyama and J.-I. Hiramatsu, *J. Biosci. Bioeng.*, 2003, **95**, 204–205.

154 R. Rowlett and J. O. E. Murphy, *Anal. Biochem.*, 1981, **112**, 163–169.

155 S. L. Zandy, J. M. Doherty, N. D. Wibisono and R. A. Gonzales, *J. Chromatogr. B*, 2017, **1055**–**1056**, 1–7.

156 C. Yang, X. Jiang, L. Guo, H. Zhang and M. Liu, *J. Sep. Sci.*, 2007, **30**, 3154–3163.

157 R. Bhushan and H. Bru, *Amino Acids*, 2004, **27**, 231–247.

158 N. J. Ayon, A. D. Sharma and W. G. Gutheil, *J. Am. Soc. Mass Spectrom.*, 2019, **30**, 448–458.

159 R. Li, H. J. Wijma, L. Song, Y. Cui, M. Otzen, Y. Tian, J. Du, T. Li, D. Niu, Y. Chen, J. Feng, J. Han, H. Chen, Y. Tao, D. B. Janssen and B. Wu, *Nat. Chem. Biol.*, 2018, **14**, 664–670.

160 K. Fujii, Y. Ikai, T. Mayumi, H. Oka and M. Suzuki, *Anal. Chem.*, 1997, **69**, 3346–3352.

161 A. M. Tomlins, P. J. D. Foxall, M. J. Lynch, J. Parkinson, J. R. Everett and J. K. Nicholson, *Biochim. Biophys. Acta - Gen. Subj.*, 1998, **1379**, 367–380.

162 H. K. Weber, H. Weber and R. J. Kazlauskas, *Tetrahedron: Asymmetry*, 1999, **10**, 2635–2638.

163 M. T. Reetz, A. Eipper, P. Tielmann and R. Mynott, *Adv. Synth. Catal.*, 2002, **344**, 1008–1016.

164 K. F. Biegasiewicz, S. J. Cooper, X. Gao, D. G. Oblinsky, J. H. Kim, S. E. Garfinkle, L. A. Joyce, B. A. Sandoval, G. D. Scholes and T. K. Hyster, *Nat. Sci.*, 2019, **364**, 1166–1169.

165 J. S. Laursen, J. Engel-Andreasen, P. Fristrup, P. Harris and C. A. Olsen, *J. Am. Chem. Soc.*, 2013, **135**, 2835–2844.

166 M. Källberg, H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu and J. Xu, *Nat. Potocol*, 2012, **7**, 1511–1522.

167 B. P. Nocek, D. M. Gillner, Y. Fan, R. C. Holz and A. Joachimiak, *J. Mol. Biol.*, 2010, **397**, 617–626.

168 W. C. McGregor, D. M. Gillner, S. I. Swierczek, D. Liu and R. C. Holz, *Springerplus*, 2013, **2**, 1–11.

169 B. Nocek, C. Reidl, A. Starus, T. Heath, D. Bienvenue, J. Osipiuk, R. Jedrzejczak, A. Joachimiak, D. P. Becker and R. C. Holz, *Biochemistry*, 2018, **57**, 574–584.

170 H. Liu and J. H. Naismith, *BMC Biothechnology*, 2008, **8**, 91.

171 K. A. Mccall, C. Huang and C. A. Fierke, *J. Nutr.*, 2000, **130**, 1437S-1446S.

172 B. L. Vallee and D. S. Auld, *Biochemistry*, 1993, **32**, 6493–6500.

173 D. W. Christianson and J. D. Cox, *Annu. Rev. Biochem.*, 1999, **68**, 33–57.

174 M. Sabaty, S. Grosse, G. Adryanczyk, S. Boiry, F. Biaso, P. Arnoux and D. Pignol, *BMC Biochem.*, 2013, **14**, 28.

175 K. A. Majorek, M. L. Kuhn, M. Chruszcz, W. F. Anderson and W. Minor, *Protein Sci.*, 2014, **23**, 1359–1368.

176 W. T. Booth, C. R. Schlachter, S. Pote, N. Ussin, N. J. Mank, V. Klapper, R. O. Lesa, C. Tang, B. K. Hurlburt and M. Chruszcz, *ACS Omega*, 2018, **3**, 760–768.

177 M. T. Reetz, *Angew. Chem. Int. Ed.*, 2011, **50**, 138–174.

178 X. Kang, X. Zhang, L. Hong, F. Peng, Z. Liu and Y. Zheng, *Process Biochem.*, 2019, **76**, 136–141.

179 M. C. Garcia Alvarez-Coque, M. J. Medina Hernandez, R. M. Villanueva Camanas and C. Mongay Fernandez, *Anal. Biochem.*, 1989, **178**, 1–7.

180 K. Miyazaki, *MEGAWHOP Cloning : A Method of Creating Random Mutagenesis Libraries via Megaprimer PCR of Whole Plasmids*, Elsevier Inc., 1st edn., 2011, vol. 498.

181 K. Miyazaki and M. Takenouchi, *Bio Tech.*, 2002, **22**, 1033–1038.

182 D. G. Gibson, L. Young, R. Chuang, J. C. Venter, C. A. H. Iii, H. O. Smith and N. America, *Nat. Methods*, 2009, **6**, 343–347.

183 O. Hara, T. Murakami, S. Imai, H. Anzai, R. Itoh, Y. Kumada, E. Takano, E. Satoh, A. Satoh, K. Nagaoka and C. Thompson, *J. Gen. Microbiol.*, 1991, **137**, 351–359.

184 S. Prasad, P. B. Khadatare and I. Roy, *Appl Env. Microbiol*, 2011, **77**, 4603–4609.

185 M. Schmidt and U. T. Bornscheuer, *Biomol. Eng.*, 2005, **22**, 51–56.

186 X. Robert and P. Gouet, *Nucleic Acids Res.*, 2014, **42**, W320–W324.

187 F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn and R. Lopez, *Nucleic Acids Res.*, 2019, **47**, W636—W641.

188 G. J. Palm, E. Fernández-álvaro, X. Bogdanovi, S. Bartsch, J. Sczodrok, R. K. Singh, D. Böttcher and H. Atomi, *Appl Microbiol Biotechnol*, 2011, **91**, 1061–1072.

189 H. Su, B. Zhou, Y. Zhang and X. Zhang, *J. Biol. Chem.*, 2015, **290**, 11188–11198.

190 M. Schweiger, R. Schreiber, G. Haemmerle, A. Lass, C. Fledelius, P. Jacobsen, H. Tornqvist, R. Zechner and R. Zimmermann, *J. Biol. Chem.*, 2006, **281**, 40236–40241.

191 D. D. C. Dr. Guiomar Sánchez Carrón, Dr. Toni Fleming, Dr. Gideon Grogan, Dr. Karen E. Holt-Tiffin, *" Exploring , evolving and exploiting coupled racemase / acylase biotransformation systems "*, 2015.

192 S. Baxter, " *Directed evolution of an industrial N -acetyl-amino acid racemase* ", 2013.

193 S. Tokuyama, *J. Mol. Catal. - B Enzym.*, 2001, **12**, 3–14.

194 S. L. Lovelock, R. C. Lloyd and N. J. Turner, *Angew. Chem. Int. Ed.*, 2014, **53**, 4652–4656.

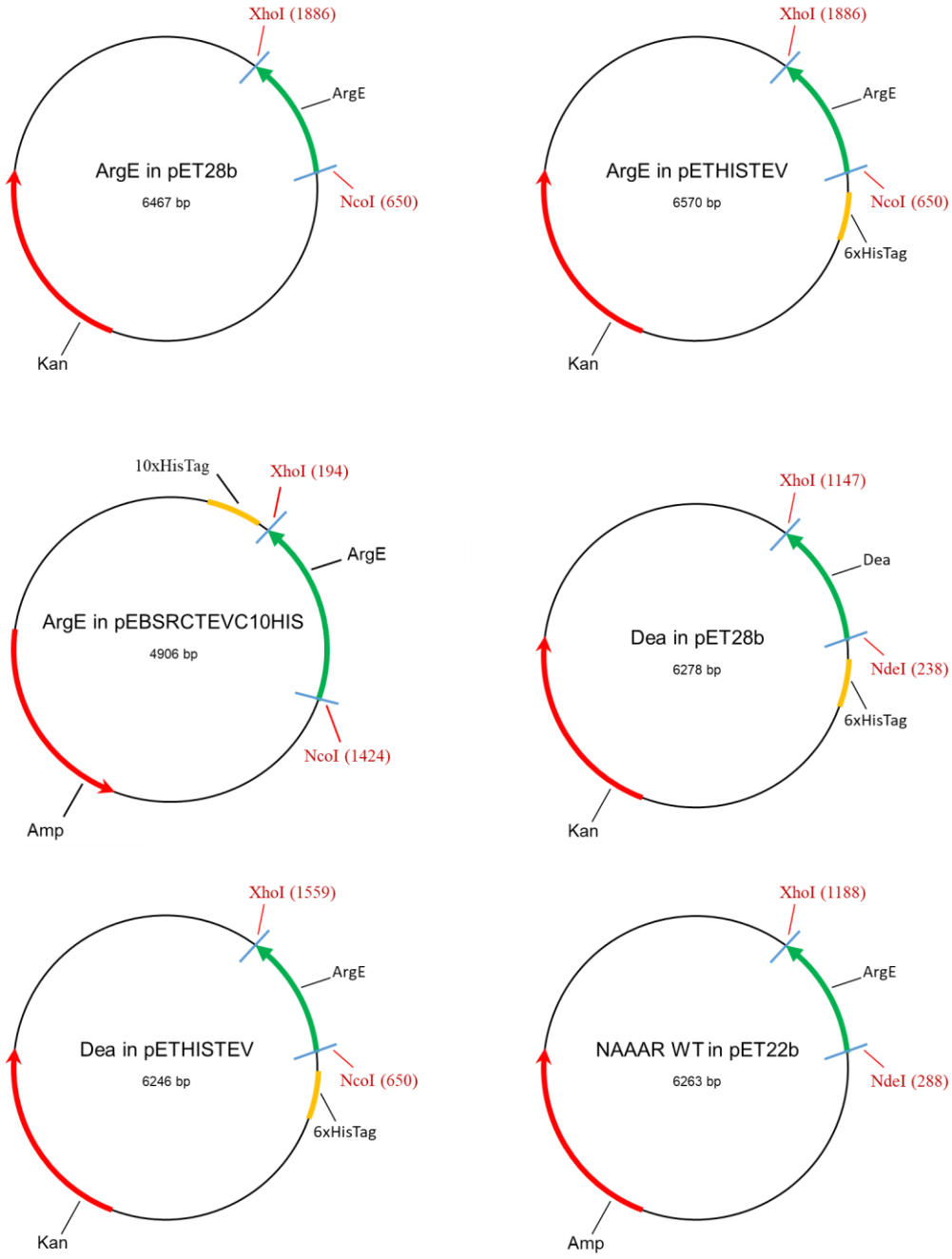
195 W. Wang, H. Xi, Q. Bi, Y. Hu, Y. Zhang and M. Ni, *Microbiol. Res.*, 2013, **168**, 360–366.

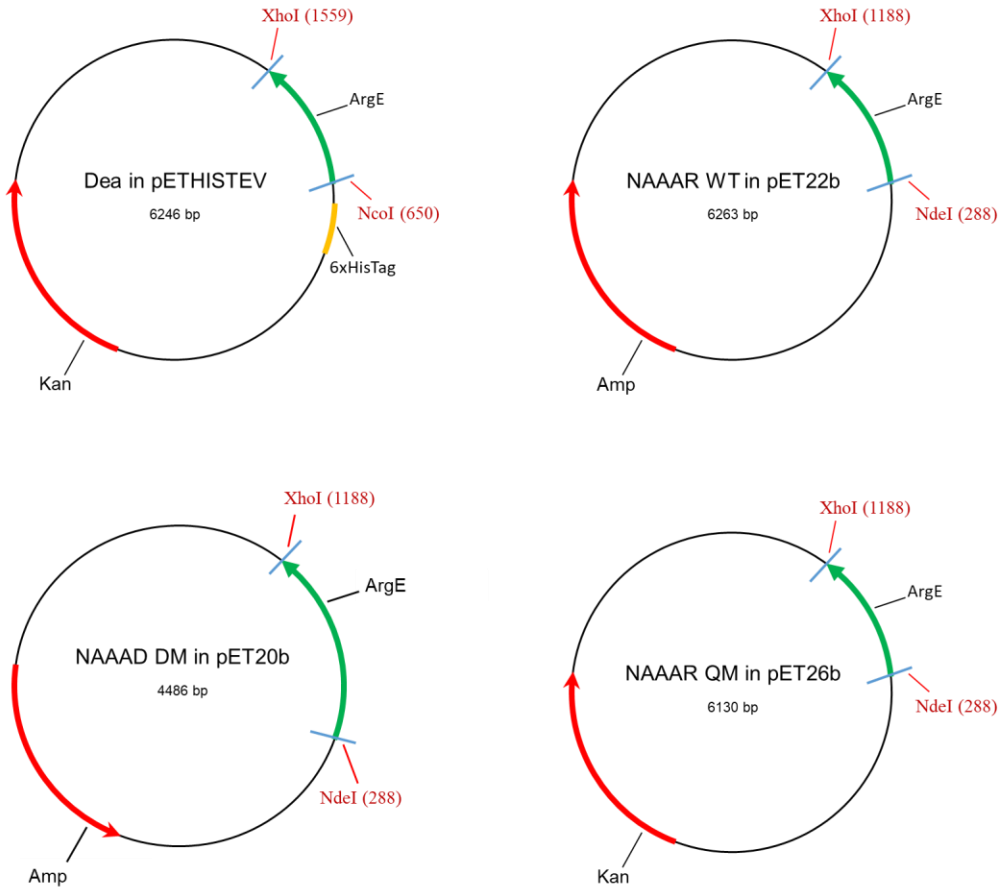
196 L. Wang, L. Wei, Y. Chen and R. Jiang, *J. Biotechnol.*, 2010, **150**, 57–63.

197 M. Talebi, S. Vaezifar, F. Jafary, M. Fazilati and S. Motamedi, *Iran J Biotech*, 2016, **14**, 33–38.

9 Appendices

Appendix 1 Plasmid map of all expression constructs used in this work.





Appendix 2 *E. coli* ArgE gene sequence (Uniprot accession code P23908).

ATGAAAAACAAATTACCGCCATTTATCGAGATTTACCGCGCTCTGATTGCCACACCTTCAAT
AAGCGCCACGGAAGAGGCACTCGATCAAAGCAATGCAGATTTAATCACTCTGCTGGCGGACT
GGTTTAAAGATTTGGGCTTCAATGTGGAAGTGCAGCCTGTTCCAGGAACTCGCAACAAATTC
AATATGCTGGCAAGTATCGGACAGGGGGCTGGCGGCTTGTTGCTGGCGGGGCATACCGATAC
GGTGCCATTTGATGACGGTCGCTGGACGCGCGATCCGTTTACACTGACGGAGCATGACGGCA
AGCTTTACGGCTTAGGCACCGCCGACATGAAAGGCTTTTTTGCGTTTATCCTTGATGCGCTA
CGCGATGTCGACGTCACGAAACTGAAAAAACCGCTCTACATTCTGGCGACTGCTGATGAAGA
AACCAGTATGGCCGGAGCGCGTTATTTTGCCGAAACTACCGCCCTGCGCCCGGATTGCGCCA
TCATTGGCGAACCGACGTCACTACAACCGGTACGCGCACATAAAGGTCATATCTCTAACGCC
ATCCGTATTCAGGGCCAGTCGGGGCACTCCAGCGATCCAGCACGCGGAGTTAACGCTATCGA
ACTAATGCACGACGCCATCGGGCATATTTTGCAATTGCGCGATAACCTGAAAGAACGTTATC
ACTACGAAGCGTTTACCGTGCCATACCCTACGCTCAACCTCGGGCATATTCACGGTGGCGAC
GCTTCTAACCGTATTTGCGCTTGCTGTGAGTTGCATATGGATATTCGTCCGCTGCCTGGCAT
GACACTCAATGAACTTAATGGTTTGCTCAACGATGCATTGGCTCCGGTGAGCGAACGCTGGC
CGGGTCGTCTGACGGTCGACGAGCTGCATCCGCCGATCCCTGGCTATGAATGCCCACCGAAT

CATCAACTGGTTGAAGTGGTTGAGAAATTGCTCGGAGCAAAAACCGAAGTGGTGAACTACTG
TACCGAAG

Appendix 3 Untagged *E. coli* ArgE amino acid sequence from pET28a construct.

MKNKLPPFIEIYRALIATPSISATEEALDQSNADLITLLADWFKDLGFNVEVQPVPGTRNKF
NMLASIGQGAGGLLLAGHTDTVPFDDGRWTRDPFTLTEHDGKLYGLGTADMKGFFAFILDAL
RDVDVTKLKKPLYILATADEETSMAGARYFAETTALRPDCAIIGEPTSLQPVRAHKGHISNA
IRIQGQSGHSSDPARGVNAIELMHDAIGHILQLRDNLKERYHYEAFTVPYPTLNLGHIHGGD
ASNRICACCELHMDIRPLPGMTLNELNGLLNDALAPVSERWPGRLTVDELHPPIPGYECPPN
HQLVEVVEKLLGAKTEVVNYCTEAPFIQTLCPTLVLGPGSINQAHQPDEYLETRFIKPTREL
ITQVIHHFCWH

Appendix 4 N-His tagged *E. coli* ArgE amino acid sequence from pETHISTEV construct.

MSYYHHHHHHDYDIPTTENLYFQGAMKNKLPPFIEIYRALIATPSISATEEALDQSNADLIT
LLADWFKDLGFNVEVQPVPGTRNKFNMLASIGQGAGGLLLAGHTDTVPFDDGRWTRDPFTLT
EHDGKLYGLGTADMKGFFAFILDALRDVDVTKLKKPLYILATADEETSMAGARYFAETTALR
PDCAIIGEPTSLQPVRAHKGHISNAIRIQGQSGHSSDPARGVNAIELMHDAIGHILQLRDNL
KERYHYEAFTVPYPTLNLGHIHGGDASNRICACCELHMDIRPLPGMTLNELNGLLNDALAPV
SERWPGRLTVDELHPPIPGYECPPNHQLVEVVEKLLGAKTEVVNYCTEAPFIQTLCPTLVLG
PGSINQAHQPDEYLETRFIKPTRELITQVIHHFCWH

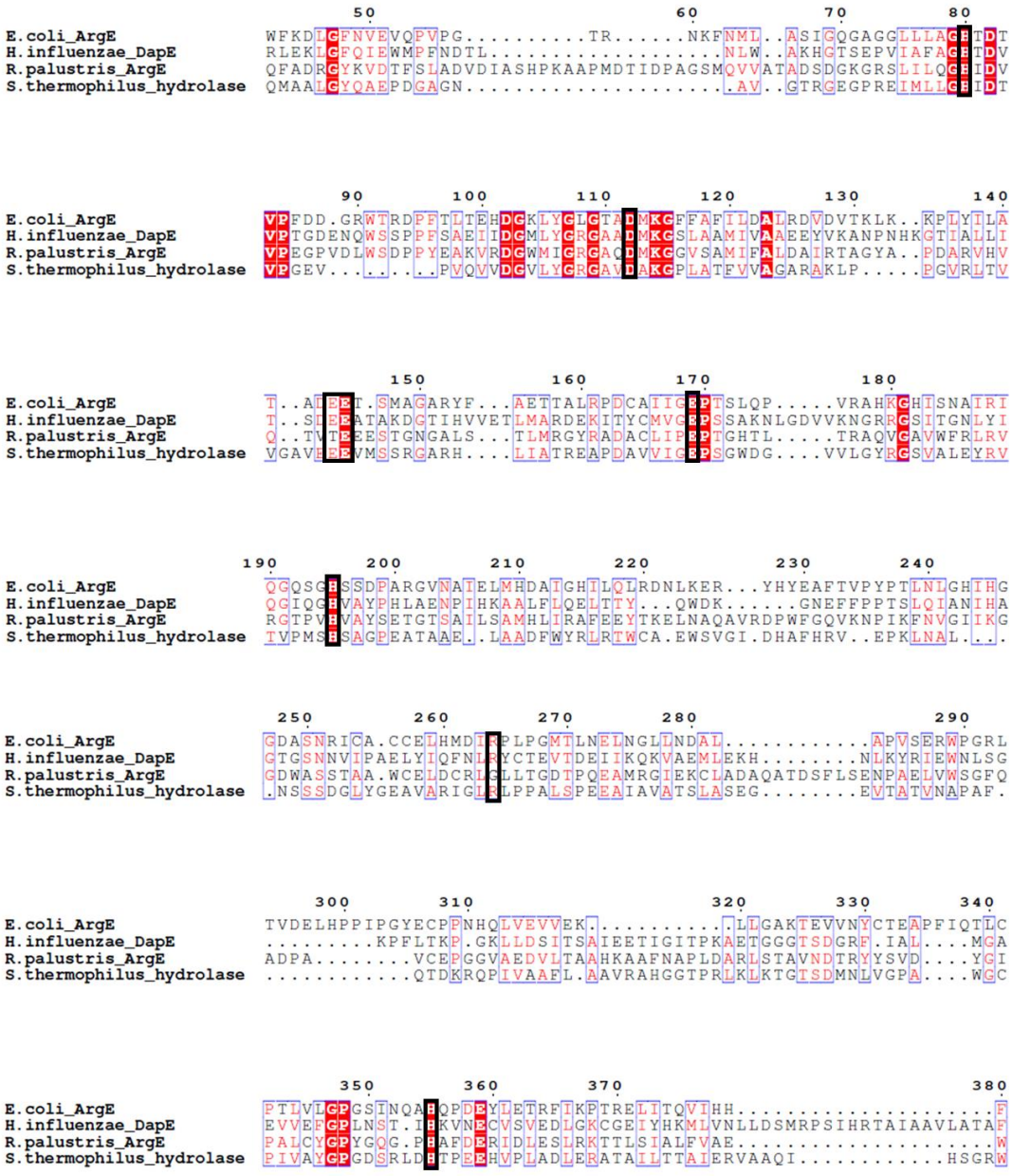
Appendix 5 C-His tagged *E. coli* ArgE amino acid sequence from pEBSRCTEVC10HIS construct.

MKNKLPPFIEIYRALIATPSISATEEALDQSNADLITLLADWFKDLGFNVEVQPVPGTRNKF
NMLASIGQGAGGLLLAGHTDTVPFDDGRWTRDPFTLTEHDGKLYGLGTADMKGFFAFILDAL
RDVDVTKLKKPLYILATADEETSMAGARYFAETTALRPDCAIIGEPTSLQPVRAHKGHISNA
IRIQGQSGHSSDPARGVNAIELMHDAIGHILQLRDNLKERYHYEAFTVPYPTLNLGHIHGGD
ASNRICACCELHMDIRPLPGMTLNELNGLLNDALAPVSERWPGRLTVDELHPPIPGYECPPN
HQLVEVVEKLLGAKTEVVNYCTEAPFIQTLCPTLVLGPGSINQAHQPDEYLETRFIKPTREL
ITQVIHHFCWH**GLEENLYFQGAHHHHHHHHHH**

Appendix 6 Table containing all the primes used for the SDM of the ArgE gene and their sequences, the mutated sites are in bold.

Primers	Sequence
ArgE E144A Fwd	GCT GAT **GCA** GAA ACC AGT ATG GCC
ArgE E144A Rev	ACT GGT TTC TGC ATC **TGC** AGT CGC
ArgE E145A Fwd	GAT GAA **GCA** ACC AGT ATG GCC GGA
ArgE E145A Rev	TCC GGC CAT ACT GGT **TGC** TTC ATC
ArgE E169A Fwd	ATT GGC **GCA** CCG ACG TCA CTA CAA
ArgE E169A Rev	TTG TAG TGA CGT CGG **TGC** GCC AAT
ArgE D112A Fwd	ATC CTT **GCT** GCG CTA CGC GAT GTC GAC
ArgE D112A Rev	TAG CGC **AGC** AAG GAT AAA CGC AAA AAA
ArgE H195I Fwd	TCG GGG **AAT** TCC AGC GAT CCA
ArgE H195I Rev	GGA **TTA** CCC CGA CTG GCC CTG AAT
ArgE R264A Fwd	GAT ATT **GCT** CCG CTG CCT GGC ATG
ArgE R264A Rev	CAT GCC AGG CAG CGG **AGC** AAT ATC
C-HisTag ArgE Fwd	TG **GCA** TGG ACT CGA GCA CCA C
C-HisTag ArgE Rev	G TGG TGC TCG AGT CCA **TGC** CA

Appendix 7 Sequence alignment between *E. coli* ArgE (Uniprot accession code P23908), DapE from *H. influenzae* (31.5 % sequence identity, PDB entry code 3IC1), *N*-acetyl-ornithine deacetylase from *Rhodopseudomonas palustris* (27.5 % sequence identity, PDB entry code 3PFO) and *N*-acetyl-ornithine hydrolase from *Spaerobacter thermophilus* (29.0 % sequence identity, PDB entry code 4Q7A) obtained from CLUSTAL omega. White over red residues are conserved across all species, red on white are conserved across most while black on white are variable across species. The residues of interest are circled in black.



Appendix 8 *S. viridochromogenes* Tü494 Dea gene sequence (Uniprot accession code Q56171).

ATGGCGAGCAGCGAACTGGAACTGGTGCGTGAACTGATTGGTCTGAACTGGCATACCCGTAA
TGGCGATGTGGAGCCGCGTCGTGTGGCGTACGACCGTGCGCAGGAAGCGTTTGGTGATCTGG
GCCTGCCGCCGGGTGAAACCGTGGTTATCGGCGACTGCAGCGCGGAGTGGGTGCGTCCGGCG
CAGGAAGATGGTCGTACCCTGCTGTACCTGCACGGTGGCAGCTATGCGCTGGGCAGCCCGCA
AAGCCACCGTCACCTGAGCAGCGCGCTGGGTGCGGCGGCGGGTGCGGCGGTTCTGGCGCTGC
ACTACCGTCGTCCGCCGGAGAGCCCGTTCCCGGCGGCGGTGGAAGACGCGGTTGCGGCGTAT
CGTATGCTGCGTGAGCGTGGTTGCCCGCCGGGTCGTATTACCTTCGCGGGTGATAGCGCGGG
TGCGGGCCTGGCGGTTGCGGCGCTGCAAGTTCTGCGTGATGCGGGCGATCCGCTGCCGGCGG
CGGCGGTGTGCATTAGCCCGTGGGCGGACCTGGCGTGCGAGGGTGCGAGCCATGTTACCCGT
AAGGAGCGTGAAATTCTGCTGGACACCGAAGATCTGCTGCGTATGGCGGGTCGTTATCTGGC
GGGTACCGACCCGCGTAACCCGCTGGCGAGCCCGGCGCACGGTGATCTGACCGGCCTGCCGC
CGCTGCTGATCCAAGTGGGTAGCGAGGAAGTTCTGTATGATGATGCGCGTGCGCTGGAGCAA
GCGGCGCTGAAAGCGGGCGTGCCGGTTACCTTCGATGAGTGGCCGGAAATGTTTCACGTGTG
GCACTGGTATCACCCGGTTCTGCCGGAAGGTCGTGCGGCGGTTGAGACCGCGGGTGTGTTTC
TGCGTCGTGCGACCGAAGAGGGCGAACGCTAA

Appendix 9 N-His tagged *S. viridochromogenes* Tü494 Dea amino acid sequence from pET28a construct.

MGSSHHHHHHSSGLVPRGSHMASSELELVRELIGLNWHTRNGDVEPRRVAYDRAQEAFGDLG
LPPGETVVIGDCSAEWVRPAQEDGRTLLYLHGGSYALGSPQSHRHLSSALGAAAGAAVLALH
YRRPPESPFPAAVEDAVAAYRMLRERGCPPGRITFAGDSAGAGLAVAALQVLRDAGDPLPAA
AVCISPWADLACEGASHVTRKEREILLDTEDLLRMAGRYLAGTDPRNPLASPAHGDLTGLPP
LLIQVGSEEVLYDDARALEQAALKAGVPVTFDEWPEMFHVWHWYHPVLPEGRAAVETAGVFL
RRATEEGER

Appendix 10 N-His tagged *S. viridochromogenes* Tü494 Dea amino acid sequence from pETHISTEV construct.

MSYYHHHHHHDYDIPTTENLYFQGAMASSELELVRELIGLNWHTRNGDVEPRRVAYDRAQEA
FGDLGLPPGETVVIGDCSAEWVRPAQEDGRTLLYLHGGSYALGSPQSHRHLSSALGAAAGAA
VLALHYRRPPESPFPAAVEDAVAAYRMLRERGCPPGRITFAGDSAGAGLAVAALQVLRDAGD
PLPAAAVCISPWADLACEGASHVTRKEREILLDTEDLLRMAGRYLAGTDPRNPLASPAHGDL
TGLPPLLIQVGSEEVLYDDARALEQAALKAGVPVTFDEWPEMFHVWHWYHPVLPEGRAAVET
AGVFLRRATEEGER

Appendix 11 *Amycolaptopsis* sp. TS1-60A NAAAR WT gene sequence

ATGGCGAGCAGCGAACTGGAACTGGTGCGTGAACTGATTGGTCTGAACTGGCATACCCGTAATGG
CGATGTGGAGCCGCGTCGTGTGGCGTACGACCGTGCGCAGGAAGCGTTTGGTGATCTGGGCCTGC
CGCCGGGTGAAACCGTGGTTATCGGCGACTGCAGCGCGGAGTGGGTGCGTCCGGCGCAGGAAGA
TGGTCGTACCCTGCTGTACCTGCACGGTGGCAGCTATGCGCTGGGCAGCCCGCAAAGCCACCGTCA
CCTGAGCAGCGCGCTGGGTGCGGCGGCGGGTGCGGCGGTTCTGGCGCTGCACTACCGTCGTCCGC
CGGAGAGCCCGTTCCCGGCGGCGGTGGAAGACGCGGTTGCGGCGTATCGTATGCTGCGTGAGCGT
GGTTGCCCGCCGGGTCGTATTACCTTCGCGGGTGATAGCGCGGGTGCGGGCCTGGCGGTTGCGGC
GCTGCAAGTTCTGCGTGATGCGGGCGATCCGCTGCCGGCGGCGGCGGTGTGCATTAGCCCGTGGG
CGGACCTGGCGTGCGAGGGTGCGAGCCATGTTACCCGTAAGGAGCGTGAAATTCTGCTGGACACC
GAAGATCTGCTGCGTATGGCGGGTCGTTATCTGGCGGGTACCGACCCGCGTAACCCGCTGGCGAG
CCCGGCGCACGGTGATCTGACCGGCCTGCCGCCGCTGCTGATCCAAGTGGGTAGCGAGGAAGTTCT
GTATGATGATGCGCGTGCGCTGGAGCAAGCGGCGCTGAAAGCGGGCGTGCCGGTTACCTTCGATG
AGTGGCCGGAAATGTTTCACGTGTGGCACTGGTATCACCCGGTTCTGCCGGAAGGTCGTGCGGCG
GTTGAGACCGCGGGTGTGTTTCTGCGTCGTGCGACCGAAGAGGGCGAACGCTAA

Appendix 12 untagged *Amycolaptopsis* sp. TS1-60A NAAAR WT amino acid sequence from pET26b construct.

MKLSGVELRRVQMPLVAPFRTSFGTQSVRELLLLRAVTPAGEGWGECVTMAGPLYSSEYNDG
AEHVLRHYLIPALLAAEDITAAKVTPLLAKFKGHRMAKGALEMAVLDAELRAHERSFAAELG
SVRDSVPCGVSVGIMDTIPQLLDVVGGYLDEGYVRIKLKIEPGWDVEPVRAVRERFGDDVLL
QVDANTAYTLGDAPQLARLDPFGLLLIEQPLEEEDVLGHAELARRIQTPICLDESIVSARAA
ADAIKLGAVQIVNIKPGRVGGYLEARRVHDVCAAHGIPVWCGGMIETGLGRAANVALASLPN
FTLPGDTSASDRFYKTDITEPFVLSGGHLPVPTGPGLGVAPIPELLDEVTTAKVWIGS

Appendix 13 untagged *Amycolaptopsis* sp. TS1-60A NAAAR DM amino acid sequence from pET26b construct. Mutations reported in red.

MKLSGVELRRVQMPLVAPFRTSFGTQSVRELLLLRAVTPAGEGWGECVTMAGPLYSSEYNDG
AEHVLRHYLIPALLAAEDITAAKVTPLLAKFKGHRMAKGALEMAVLDAELRAHERSFAAELG
SVRDSVPCGVSVGIMDTIPQLLDVVGGYLDEGYVRIKLKIEPGWDVEPVRAVRERFGDDVLL
QVDANTAYTLGDAPQLARLDPFGLLLIEQPLEEEDVLGHAELARRIQTPICLDESIVSARAA
ADAIKLGAVQIVNIKPGRVGGYLEARRVHDVCAAHGIPVWCG**D**MIETGLGRAANVALASLPN
FTLPGDTSASDR**Y**YKTDITEPFVLSGGHLPVPTGPGLGVAPIPELLDEVTTAKVWIGS

Appendix 14 untagged *Amycolaptopsis* sp. TS1-60A NAAAR QM amino acid sequence from pET26b construct. Mutations reported in red.

MKLSGVELRRVQMPLVAPFRTSFGT**A**SVRELLLLRAVTPAGEGWGECVT**I**AGPLYSSEYNDG
AEHVLRHYLIPALLAAEDITAAKVTPLLAKFKGHRMAKGALEMAVLDAELRAHERSFAAELG
SVRDSVPCGVSVGIMDTIPQLLDVVGGYLDEGYVRIKLKIEPGWDVEPVRAVRERFGDDVLL
QVDANTAYTLGDAPQLARLDPFGLLLIEQPLEEEDVLGHAELARRIQTPICLDESIVSARAA

ADAIKLGAVQIVNIKPGRVGGYLEARRVHDVCAAHGIPVWCG**D**MIETGLGRAANVALASLPN

FTLPGDTSASDR**Y**YKTDITEPFVLSGGHLPVPTGPGLGVAPIPELLDEVTTAKVWIGS

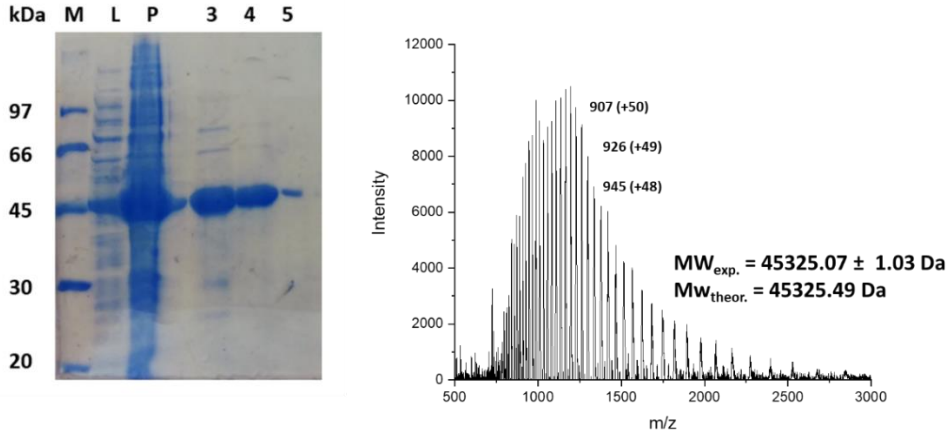
Appendix 15 X-ray data collection for *E. coli* ArgE E144A mutant crystals. Numbers in brackets refer to data for highest resolution shells.

	ArgE
Beamline	Diamond I04-1
Wavelength (Å)	0.91587
Resolution (Å)	215.55-2.75 (2.83-2.75)
Space Group	$P6_322$
Unit cell (Å)	a = b = 248.90 ; c = 116.74
	α = β = 90.0 γ = 120.0 °
No. of molecules in the asymmetric unit	4
Unique reflections	55565 (4487)
Completeness (%)	100.0 (100.0)
R_{merge} (%)	0.38 (4.95)
$R_{p.i.m.}$	0.11 (1.50)
Multiplicity	23.1 (23.2)
$\langle I/\sigma(I)\rangle$	8.5 (0.9)
Overall *B* factor from Wilson plot (Å^2)	61
$CC_{1/2}$	1.00 (0.52)

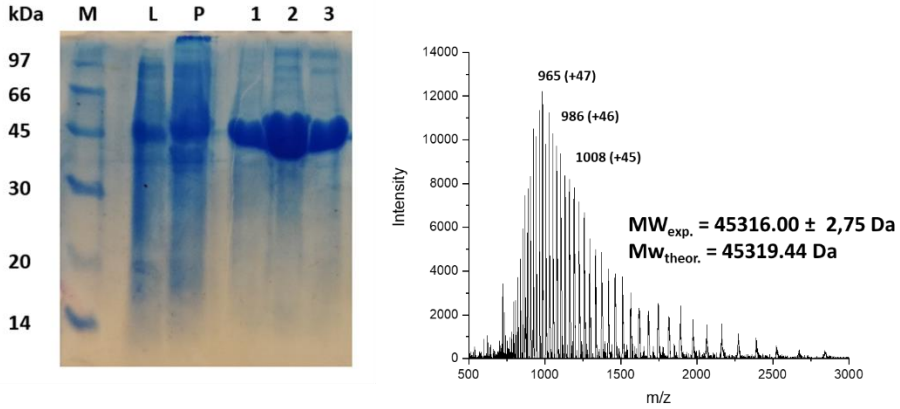
Appendix 16 X-ray data collection for *N*-acetyl phosphinothricin. Single colourless plate-shaped crystals were recrystallised from a mixture of acetone and water by slow evaporation. A suitable crystal 0.44×0.11×0.04 mm^3 was selected and mounted on a MITIGEN holder in Paratone oil on an Rigaku Oxford Diffraction SuperNova diffractometer. The crystal was kept at a steady T = 120.0 K during data collection. The structure was solved with the ShelXT (Sheldrick, 2015) structure solution program using the Intrinsic Phasing solution method and by using Olex2 (Dolomanov et al., 2009) as the graphical interface. The model was refined with version 2018/3 of ShelXL (Sheldrick, 2015) using Least Squares minimisation.

***N*-Acetyl phosphinothricin**	**DC19001**
$D_{calc.}$/ g cm^{-3}	1.411
m/mm^{-1}	2.366
Formula Weight	223.16
Colour	colourless
Shape	plate
Size/mm^3	0.44×0.11×0.04
T/K	120
Crystal System	monoclinic
Space Group	$P2_1/n$
a/Å	10.9212(5)
b/Å	6.5192(3)
c/Å	15.1831(6)
a/°	90
b/°	103.721(4)
g/°	90
V/$Å^3$	1050.15(8)
Z	4
Z'	1
Wavelength/Å	1.54184
Radiation type	CuK_a
Q_{min}/°	4.52
Q_{max}/°	76.303
Measured Refl.	16514
Independent Refl.	2184
Reflections with I > 2(I)	1972
R_{int}	0.0717
Parameters	138
Restraints	3
Largest Peak	0.552
Deepest Hole	-0.477
GooF	1.062
wR_2 (all data)	0.1472
wR_2	0.1431
R_1 (all data)	0.0556
R_1	0.052

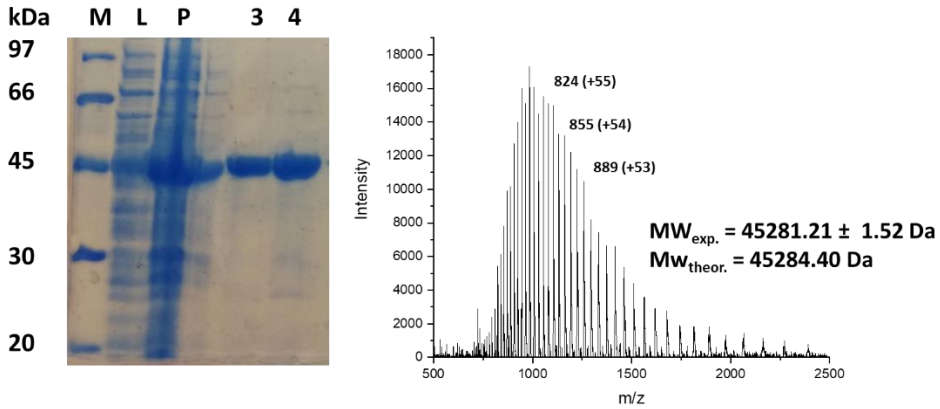
Appendix 17 ArgE D112A mutant characterization *via* SDS-PAGE gel electrophoresis and denaturing LC-ESI MS spectrometry. 12% SDS-PAGE gel, lane 1 LMWM (M), lane 2 soluble CFE (L), lane 3 cell pellets (P) and lane 4-6 HisTrap fractions (~3 mL each).



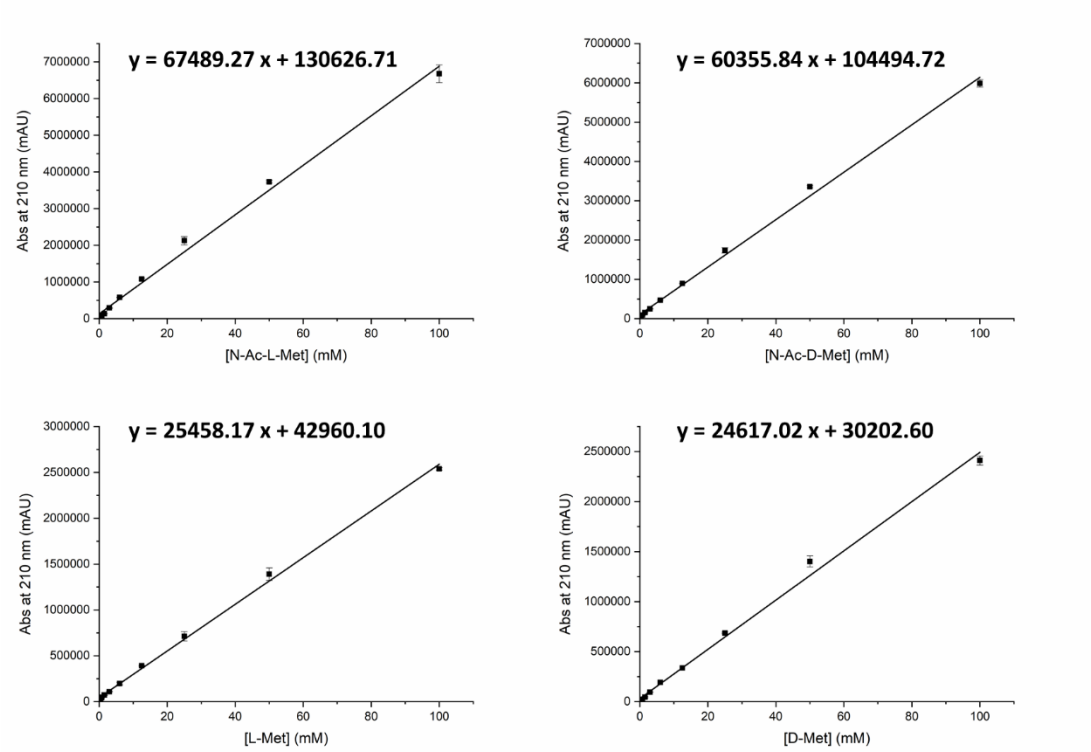
Appendix 18 ArgE H195I mutant characterization *via* SDS-PAGE gel electrophoresis and denaturing LC-ESI MS spectrometry. 12% SDS-PAGE gel, lane 1 LMWM (M), lane 2 soluble CFE (L), lane 3 cell pellets (P) and lane 4-6 HisTrap fractions (~3 mL each).



Appendix 19 ArgE R264A mutant characterization *via* SDS-PAGE gel electrophoresis and denaturing LC-ESI MS spectrometry. 12% SDS-PAGE gel, lane 1 LMWM (M), lane 2 soluble CFE (L), lane 3 cell pellets (P) and lane 4, 5 HisTrap fractions (~3 mL each).

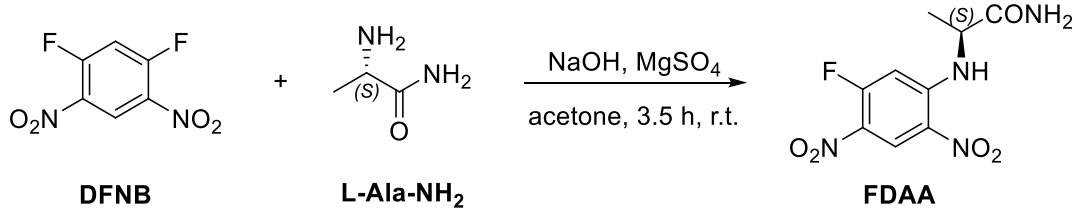


Appendix 20 Calibration curves for *N*-Ac-L-Met, *N*-Ac-D-Met, L-Met and D-Met obtained by chiral HPLC analysis. For each compound a stock solution at 100 mM concentration was prepared and diluted to obtain ten standards for the calibration curve. 20 µL of standard were diluted in 0.025% triethyl ammonium acetate (TEAA):MeOH (75:25) to prepare the samples for analysis. Samples were then run by HPLC using a Chirobiotic T column (Astec, chiral phase Teicoplanin, 5 µm, 250 mm x 4.6 mm) and the following isocratic conditions: mobile phase: 0.025% TEAA:MeOH (75:25, v/v), flow rate: 0.5 mL min-1, detection λ: 210 nm, temperature: 40 °C, run time: 30 minutes. Each sample prepared in triplicates.



Appendix 21 Characterization of reaction products isolated. NMR run on a Bruker AVA500 or Pro500 (500 MHz) instrument, with the solvent peak used as reference. LCMS samples prepared in MS grade water or methanol were run on a Bruker MicrOTOF system with an ESI source. Chiral HPLC analysis were run on a Chirobiotic T column (Astec, chiral phase Teicoplanin, 5 µm, 250 mm x 4.6 mm) and the following isocratic conditions: mobile phase: 0.025% TEAA:MeOH (40:60, v/v), flow rate: 0.5 mL min^{-1}, detection λ: 210 nm, temperature: 40 °C, run time: 30 minutes.

Marfey's Reagent (FDAA) synthesis



DFNB **L-Ala-NH_2** **FDAA**

L-alanine amide (0.131 g, 1.5 mmol) were dissolved into acetone (0.1 M); 1 mL of NaOH (1 M) was added to the mixture, which was left stirring for 5 min or until the starting material is completely dissolved. $MgSO_4$ (2.8 g) was added to the solution, which was stirred at r.t. for 3 h, before being filtered. The filtered solution was added dropwise into a second solution of di-nitro-di-fluoro benzene (173 mg, 0.85 mmol) in acetone (0.2 M). The reaction mixture was stirred at r.t. for 30 minutes and quenched by addition of cold water (20 mL) dropwise. The bright yellow needle shaped crystals formed are filtered under *vacuum*, washed with cold water, dried and stored in darkness.

Yield: 88 %

^{1}H NMR: (500 MHz, MeOD) δ 9.11 (d, *J* = 6.8 Hz, 1H), 6.96 (d, *J* = 14.3 Hz, 1H), 4.40 (p, *J* = 6.7 Hz, 1H) and 1.46 (d, *J* = 6.7 Hz, 3H).

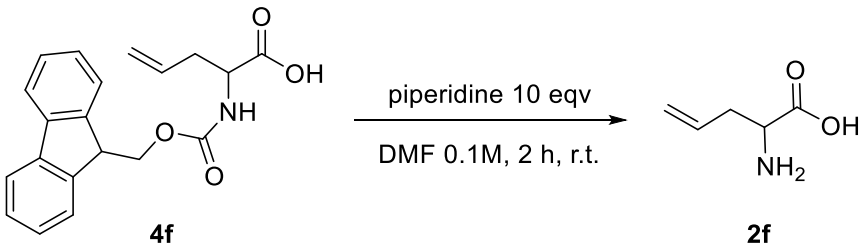
^{13}C NMR: (126 MHz, MeOD) δ 174.40, 160.79, 158.66, 147.90, 127.11, 101.47, 52.14 and 17.80.

^{19}F NMR: (500 MHz, MeOD) δ -104.22 (d, 1F)

TOF-ESI MS: calculated. 273.2 $[M + H]^+$, found 273

Commercially available.

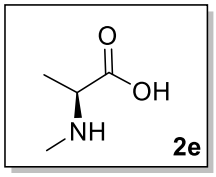
FMOC-2-amino acids de-protection



4f **2f**

FMOC-allylglycine (3g, 7.6 mmol) was placed in a round bottom flask and dissolved in DMF (0.1 M, 76 mL). Following the addition of piperidine (10 eqv, 7.5 mL) the solution was left stirring at r.t. for 2 h. The reaction was quenched by removing the solvent *in vacuo* and the residue was dissolved in hexane with the formation of a white precipitate. This precipitate was collected and washed first with hexane

(20 mL x 3) and then with EtOH (20 mL x 3). The combined ethanolic phases were concentrated under *vacuum* to obtain pure allylglycine (**2b**) as a white powder.

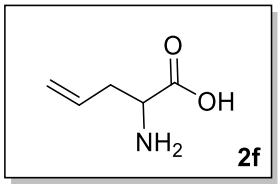


Yield: 88 %

^{1}H NMR: (500 MHz, D_2O) δ 3.50 (q, *J* = 7.2 Hz, 3H), 2.59 (s, 1H), 1.36 (d, *J* = 7.2 Hz, 3H).

^{13}C NMR: (126 MHz, D_2O) δ 174.93, 58.90, 30.98 and 14.56.

TOF-ESI MS: calculated. 114.1 $[M + H]^+$, found 114



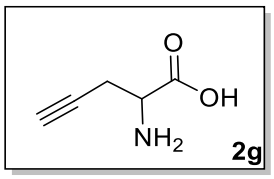
Yield: 95 %

^{1}H NMR: (500 MHz, D_2O) δ 5.67 (ddt, *J* = 17.4, 10.2, 7.3 Hz, 1H), 5.23 – 5.09 (m, 2H), 3.70 (dd, *J* = 7.1, 4.9 Hz, 1H), 2.62 – 2.44 (m, 2H).

^{13}C NMR: (126 MHz, D_2O) δ 182.74, 134.43, 117.92, 55.41 and 39.10.

TOF-ESI MS: calculated. 116.2 $[M + H]^+$, found 116

Commercially available.



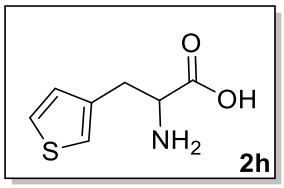
Yield: 87 %

^{1}H NMR: (500 MHz, D_2O) δ 3.81 (t, *J* = 5.5 Hz, 1H), 2.75 (ddd, *J* = 5.2, 2.7, 1.4 Hz, 2H), 2.41 (t, *J* = 2.7 Hz, 1H).

^{13}C NMR: (126 MHz, D_2O) δ 181.10, 81.32, 71.42, 54.41 and 24.42.

TOF-ESI MS: calculated 114.2 $[M + H]^+$, found 114

Commercially available.



Yield: 81 %

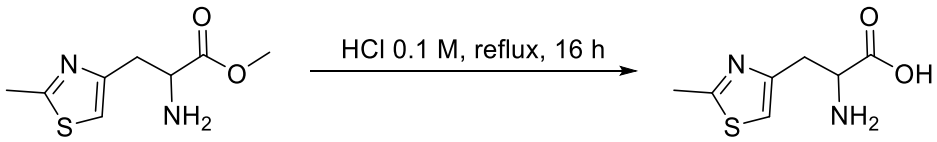
^{1}H NMR: (500 MHz, D_2O) δ 7.21 (dd, *J* = 5.1, 1.2 Hz, 1H), 6.92 (dd, *J* = 5.2, 3.4 Hz, 1H), 6.84 (dt, *J* = 3.4, 1.0 Hz, 1H), 3.48 (dd, *J* = 6.3, 5.6 Hz, 1H), 3.11 (dt, *J* = 6.2, 0.8 Hz, 2H).

^{13}C NMR: (126 MHz, D_2O) δ 181.85, 140.06, 127.23, 126.48, 124.69, 57.26 and 34.75.

TOF-ESI MS: calculated 172.1 $[M + H]^+$, found 172

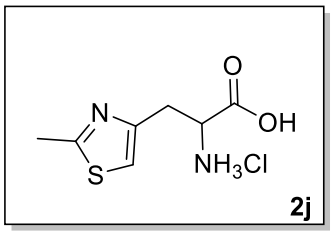
Chiral HPLC: R.T. = 9.33 min (L) and 10.20 min (D), racemate

Esters hydrolysis



5j **2j**

5j (316 mg, 1.58 mmol) was placed in a round bottom flask and dissolved in HCl 6M (0.1 M, 16 mL); this solution was stirred under reflux for 16 h. The reaction was quenched by removing the solvent under *vacuum*, isolating 2-amino-3-(2-methylthiazol-4-yl) propanoic acid hydrochloride (**2j**) as a white powder.

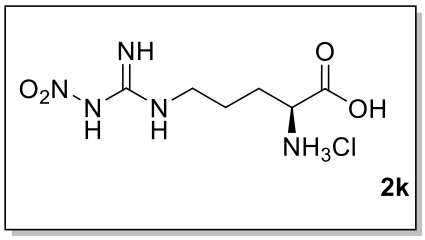


Yield: 95 %

^{1}H NMR: (500 MHz, D_2O) δ δ 7.49 (s, 1H), 4.24 – 4.19 (m, 1H), 3.63 – 3.32 (m, 2H), 2.81 (s, 3H).

TOF-ESI MS: calculated 187.1 $[M + H]^+$, found 187

Chiral HPLC: R.T. = 11.17 min (L) and 13.89 min (D), racemate

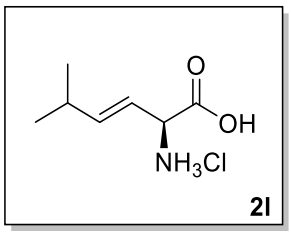


Yield: 98 %.

^{1}H NMR: (500 MHz, D_2O) δ 4.05 (t, *J* = 6.3 Hz, 1H), 3.31 (t, *J* = 6.7 Hz, 2H), 2.16 – 1.86 (m, 2H), 1.74 (dddt, *J* = 18.9, 12.5, 9.5, 6.0 Hz, 2H).

^{13}C NMR: (126 MHz, D_2O) δ 171.76, 158.92, 53.62, 52.47, 26.91 and 23.26.

TOF-ESI MS: calculated 220.2 $[M + H]^+$, found 220



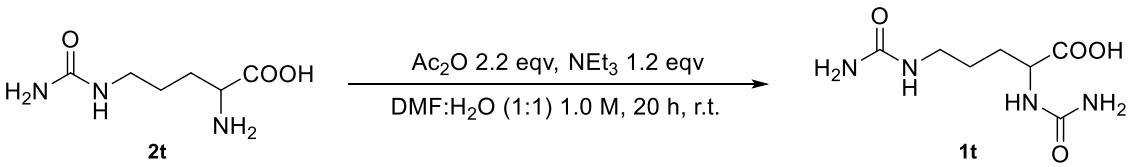
Yield: 70 %

^{1}H NMR: (500 MHz, D_2O) δ 6.02 (ddd, *J* = 15.5, 6.5, 1.0 Hz, 1H), 5.51 (ddd, *J* = 15.6, 8.2, 1.5 Hz, 1H), 4.43 (d, *J* = 8.2 Hz, 1H), 2.45 – 2.24 (m, 1H), 0.97 (d, *J* = 6.8 Hz, 6H).

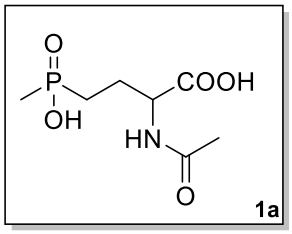
^{13}C NMR: (126 MHz, D_2O) δ 181.89, 139.55, 126.33, 58.40, 30.18, 28.48 and 21.43

TOF-ESI MS: calculated 144.1 $[M + H]^+$, found 144

Acetylation of polar amino acids



In a round bottom flask **2u** (297mg, 1.71 mmol) was dissolved in a 1:1 mixture of DMF and water (1.0 M, 1.86 mL). NEt_3 (290 µL, 1.2 eqv) and Ac_2O (317 µL, 2.2 eqv) were added dropwise to the solution under stirring. After 20 h at r.t. the solvent was removed under vacuum and the residues re-dissolved in water; the pH of the solution was adjusted to 10 with $NaCO_{3\ sat}$ and the mixture washed with Et_2O (20 mLx3). The pH of the water phase was adjusted to 2 with HCl 2 M and the solvent removed under vacuum. The white solid was dissolved in the minimum amount of MeOH, the solution filtered and methanol removed *in vacuo* to give pure *N*-acetyl-citrulline (**1t**).



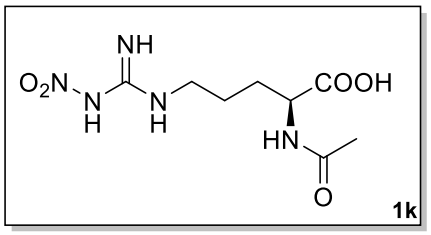
Yield: 73 %

^{1}H NMR: (500 MHz, D_2O) δ 4.40 – 4.29 (m, 1H), 2.16 – 2.02 (m, 1H), 1.96 (s, 3H), 1.94 – 1.71 (m, 3H), 1.45 (d, *J* = 14.0 Hz, 3H).

^{13}C NMR: (126 MHz, D_2O) δ 174.80, 174.36, 53.03, 52.90, 26.18, 25.44, 23.44, 23.42, 21.60, 13.16 and 13.89.

^{31}P NMR: (500 MHz, D_2O) δ 39.38 (ddq, *J* = 21.1, 14.0, 7.4 Hz).

TOF-ESI MS: calculated 222.1 $[M - H]^-$, found 222

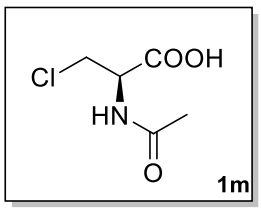


Yield: 41.4 %

^{1}H NMR: (500 MHz, D_2O) δ 4.35 (dd, *J* = 8.6, 5.0 Hz, 1H), 3.29 (t, *J* = 6.7 Hz, 2H), 2.02 (s, 3H), 1.97 – 1.63 (m, 4H).

^{13}C NMR: (126 MHz, D_2O) δ 178.83, 173.59, 158.85, 54.76, 40.75, 28.85, 23.91 and 21.88.

TOF-ESI MS: calculated 262.1151 $[M + H]^+$, found 262.1148

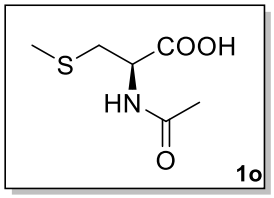


Yield: 17.8 %

^{1}H NMR: (500 MHz, D_2O) δ 4.77 (d, *J* = 7.2 Hz, 1H), 4.01 (dd, *J* = 11.7, 5.1 Hz, 1H), 3.90 (dd, *J* = 11.7, 3.6 Hz, 1H), 2.06 (s, 3H).

^{13}C NMR: (126 MHz, D_2O) δ

TOF-ESI MS: calculated 164.5 $[M - H]^-$, found 164

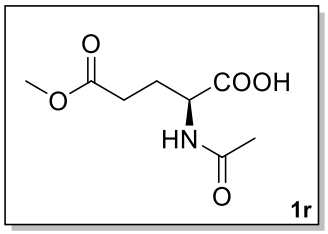


Yield: 49.6 %

^{1}H NMR: (500 MHz, D_2O) δ 4.66 – 4.60 (m, 1H), 3.10 – 3.03 (m, 1H), 2.94 (dd, *J* = 14.2, 8.3, 1H),2.96 (s, 3H), 2.08 (s, 3H).

^{13}C NMR: (126 MHz, D_2O) δ 174.30, 174.13, 52.13, 34.59, 21.64.

TOF-ESI MS: calculated $C_7H_{11}NO_3S$ 178.0551 $[M + H]^+$, found 178.0535

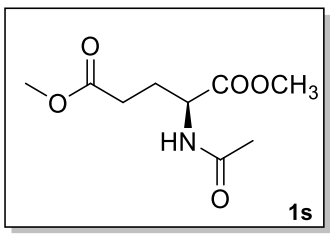


Yield: 66 %

^{1}H NMR: (500 MHz, D_2O) δ 4.38 (dd, *J* = 9.1, 5.1 Hz, 1H), 3.67 (s, 3H), 2.50 (t, *J* = 7.3 Hz, 2H), 2.28 – 2.13 (m, 1H), 2.01 (s, 3H), 2.09 – 1.89 (m, 1H).

^{13}C NMR: (126 MHz, D_2O) δ 178.31, 176.25, 173.63, 54.41, 52.23, 30.48, 26.95 and 21.90.

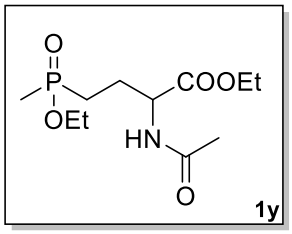
TOF-ESI MS: calculated 202.2 $[M - H]^-$, found 202



Yield: 1 %

^{1}H NMR: (500 MHz, D_2O) δ 4.42 (dd, *J* = 9.1, 5.3 Hz, 1H), 3.74 (s, 3H), 3.68 (s, 3H), 2.49 (t, *J* = 7.2 Hz, 2H), 2.20 (m, 1H), 2.01 (s, 3H), 2.06 – 1.91 (m, 1H).

TOF-ESI MS: calculated 218.2 $[M + H]^+$, found 218



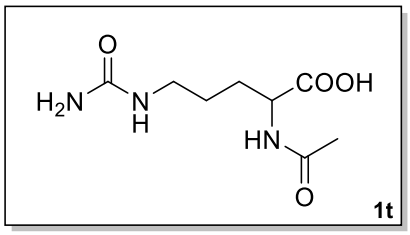
Yield: 11.5 %

^{1}H NMR: (500 MHz, D_2O) δ 4.44 (dd, *J* = 7.9, 5.0 Hz, 1H), 4.27 – 4.17 (m, 2H), 4.12 – 4.00 (m, 2H), 2.21 – 2.08 (m, 1H), 2.03 (s, 3H), 1.99 – 1.83 (m, 3H), 1.58 (d, *J* = 13.8 Hz, 3H), 1.28 (dt, *J* = 16.9, 7.1 Hz, 6H).

^{13}C NMR: (126 MHz, D_2O) δ 174.37, 173.67, 172.94, 62.82, 62.02, 61.96, 53.07, 52.94, 24.70, 23.95, 23.18, 23.11, 21.59, 15.74, 15.70, 13.27, 12.19, 11.46.

^{61}P NMR: (500 MHz, D_2O) δ 42.92 (m).

TOF-ESI MS: calculated 280.3 $[M - H]^-$, found 280



Yield: 76.3 %

^{1}H NMR: (500 MHz, D_2O) δ 4.23 – 4.15 (m, 1H), 3.09 (t, *J* = 6.8 Hz, 2H), 2.01 (s, 3H), 1.81 (m, 1H), 1.69 (m, 1H), 1.58 – 1.47 (m, 2H).

^{13}C NMR: (126 MHz, D_2O) δ 179.10, 173.57, 161.58, 55.02, 46.69, 28.90, 25.62 and 21.88.

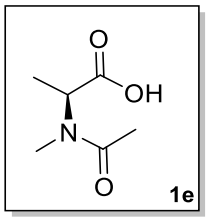
TOF-ESI MS: calculated 216.1 $[M - H]^-$, found 216

Chiral HPLC: R.T. = 4.61 min (L) and 5.46 min (D), racemate

The L-stereoisomer is commercially available.

Acetylation of hydrophobic amino acids

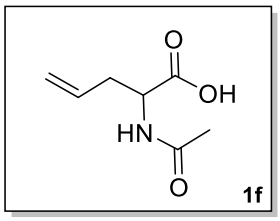
The reaction was carried out using the same protocol as **2u**. After 20 h at r.t. the solvent was removed under vacuum and the residues re-dissolved in water; the pH of the solution was adjusted to 10 with $NaCO_{3\,sat}$ and the mixture washed with Et_2O (20 mLx3). The organic phase discarded and the pH of the water phase was adjusted to 2 with HCl 2 M and the solution extracted with EtOAc (20 mLx3). The organic phase was dried under $MgSO_4$, filtered and the solvent removed under vacuum to give pure *N*-acetyl-allylglycine (**1f**).



Yield: 19 %

^{1}H NMR: (500 MHz, D_2O) δ 5.28 (s, 1H), 3.01 (s, 3H), 2.12 (s, 3H), 1.38 (s, 3H).

TOF-ESI MS: calculated 144.2 $[M - H]^-$, found 144



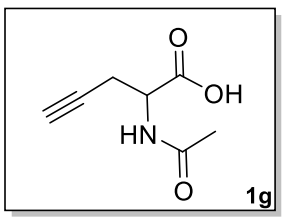
Yield: 41 %

^{1}H NMR: (500 MHz, D_2O) δ 5.78 (ddt, *J* = 17.1, 10.2, 7.0 Hz, 1H), 5.22 – 5.10 (m, 2H), 4.40 (dd, *J* = 8.2, 5.2 Hz, 1H), 2.68 – 2.42 (m, 2H), 2.01 (s, 3H).

^{13}C NMR: (126 MHz, D_2O) δ 178.48, 173.50, 133.92, 117.90, 54.83, 36.07 and 21.86.

TOF-ESI MS: calculated 156.1 $[M - H]^-$, found 156

Chiral HPLC: R.T. = 4.71 min (L) and 5.43 min (D), racemate



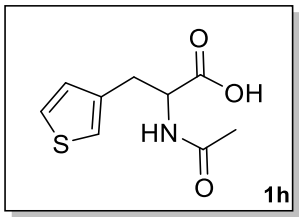
Yield: 52 %

^{1}H NMR: (500 MHz, D_2O) δ 4.53 (t, *J* = 5.9 Hz, 1H), 2.75 (ddd, *J* = 5.6, 2.7, 1.4 Hz, 2H), 2.41 (t, *J* = 2.7 Hz, 1H), 2.04 (s, 3H).

^{13}C NMR: (126 MHz, D_2O) δ 176.82, 173.64, 80.66, 71.31, 53.51, 21.90 and 21.81.

TOF-ESI MS: calculated 154.1 $[M - H]^-$, found 154

Chiral HPLC: R.T. = 4.19 min (L) and 4.85 min (D), racemate

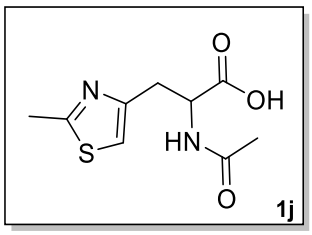


Yield: 43.0 %

^{1}H NMR: (500 MHz, D_2O) δ 7.22 (dd, *J* = 5.2, 1.2 Hz, 1H), 6.91 (dd, *J* = 5.1, 3.5 Hz, 1H), 6.85 (dd, *J* = 3.5, 1.1 Hz, 1H), 4.54 (dd, *J* = 8.3, 5.0 Hz, 1H), 3.34 (dd, *J* = 15.1, 5.0, 1H), 3.18 (dd, *J* = 15.0, 8.3, 1H), 1.88 (s, 3H).**^{13}C NMR:** (126 MHz, D_2O) δ 177.37, 173.49, 139.67, 127.11, 126.47, 124.85, 56.20, 31.80 and 21.94.

TOF-ESI MS: calculated 214.0540 $[M + H]^+$, found 214.0535

Chiral HPLC: R.T. = 4.97 min (L+D), racemate



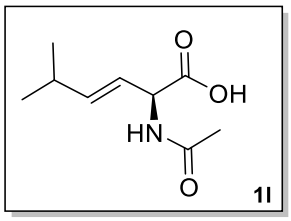
Yield: 46.7 %

^{1}H NMR: (500 MHz, D_2O) δ 7.44 (s, 1H), 4.46 (t, *J* = 6.5 Hz, 1H), 3.39 (dd, *J* = 15.1, 5.1, 1H), 3.18 (dd, *J* = 15.1, 9.0 Hz, 1H), 2.84 (s, 3H), 1.96 (s, 3H).

^{13}C NMR: (126 MHz, D_2O) δ 177.83, 173.22, 168.12, 151.23, 115.53, 55.01, 32.22, 21.84 and 17.65.

TOF-ESI MS: calculated for $C_9H_{11}ClFNO_2$ 227.1 $[M - H]^-$, found 227

Chiral HPLC: R.T. = 4.73 min (L) and 5.62 min (D), racemate

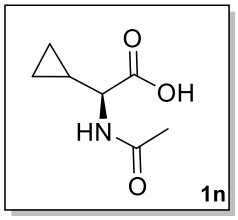


Yield: 56 %

^{1}H NMR: (500 MHz, D_2O) δ 5.99 (dd, *J* = 15.6, 6.5 Hz, 1H), 5.86 (dd, *J* = 15.5, 6.5 Hz, 1H), 5.63 (ddd, *J* = 15.6, 7.0, 4.1 Hz, 2H), 4.72 (d, *J* = 6.6 Hz, 2H), 2.15 (s, 3H), 2.15 (s, 3H), 1.08 (m, 12H).

^{13}C NMR: (126 MHz, D_2O) δ 177.69, 173.81, 173.45, 173.03, 144.22, 141.26, 122.03, 119.00, 57.48, 55.16, 53.08, 48.88, 21.92, 21.54, 21.28 and 21.03.

TOF-ESI MS: calculated 170.2 $[M - H]^-$, found 170.2

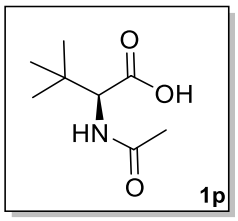


Yield: 58.8 %

^{1}H NMR: (500 MHz, D_2O) δ 3.52 (d, *J* = 9.4 Hz, 1H), 2.00 (s, 3H), 1.11 (m, 1H), 0.71 – 0.54 (m, 2H), 0.54 – 0.42 (m, 1H), 0.32 (m, 1H).

^{13}C NMR: (126 MHz, D_2O) δ 178.92, 173.37, 59.44, 21.72, 12.78, 2.92 and 2.35.

TOF-ESI MS: calculated 158.0828 $[M + H]^+$, found 158.0814

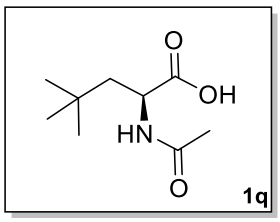


Yield: 53 %

^{1}H NMR: (500 MHz, D_2O) δ 4.11 (s, 1H), 2.04 (s, 3H) and 1.00 (s, 9H).

^{13}C NMR: (126 MHz, D_2O) δ 175.04, 174.26, 61.97, 32.99, 25.83 and 21.56.

TOF-ESI MS: calculated $C_8H_{16}O_3$ 174.1150 $[M + H]^+$, found 174.1127



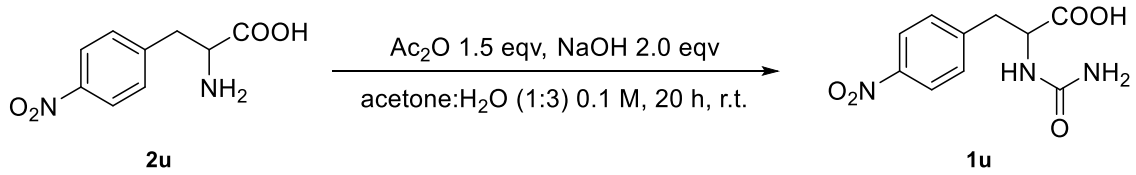
Yield: 79.3 %

^{1}H NMR: (500 MHz, D_2O) δ 4.49 (dd, *J* = 7.6, 6 Hz, 1H), 1.99 (s, 3H), 1.82 (dd, *J* = 11.6, 2.4 Hz, 1H), 1.61 (dd, *J* = 11.6, 7.6 Hz, 1H) and 1.00 (s, 9H).

^{13}C NMR: (126 MHz, D_2O) δ 180.78, 173.08, 53.16, 45.23, 29.91, 28.85 and 21.99.

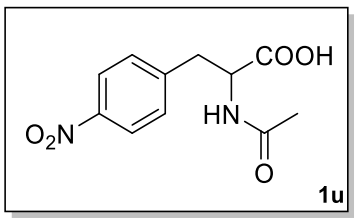
TOF-ESI MS: calculated 188.1288 $[M + H]^+$, found 188.1284

Acetylation of aromatic amino acids



2u **1u**

In a round bottom flask **2y** (1.5 g, 6.1 mmol) and NaOH (0.55 g, 2.0 eqv) were dissolved in a 1:3 mixture of acetone and water (0.1 M, 60 mL). Slowly Ac_2O (1.0 mL, 1.5 eqv) was added to the solution and after addition the mixture was left stirring for 20 h. The reaction was quenched by removal of the acetone under *vacuum* and by addition of $NaCO_{3\ sat}$ to the remaining solution until the pH reached a value of 10. The solution was then washed with Et_2O (20 mLx3) and the pH of the water phase was adjusted to 2 with HCl 2 M. This solution was extracted with EtOAc (20 mLx3) and the collected organic phase was dried under $MgSO_4$, filtered and the solvent removed under *vacuum* to give pure *N*-acetyl-4-NO_2-phenylalanine (**2u**).



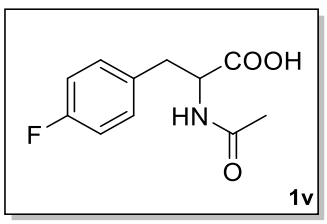
Yield: 84.0 %

^{1}H NMR: (500 MHz, D_2O) δ 8.24 (d, *J* = 8.8 Hz, 2H), 7.52 (d, *J* = 8.8 Hz, 2H), 4.76 – 4.68 (m, 1H), 3.40 (dd, *J* = 13.9, 5.3 Hz, 1H), 3.16 (dd, *J* = 14.0, 9.3 Hz, 1H), 1.96 (s, 3H).

^{13}C NMR: (126 MHz, D_2O) δ 175.00, 173.77, 146.68, 145.22,130.21, 123.68, 53.95, 36.69 and 21.50.

TOF-ESI MS: calcd. for $C_{11}H_{12}NO_5$ 252.1006 $[M - H]^-$, found 252.0998

Chiral HPLC: R.T. = 4.64 min (L) and 5.71 min (D), racemate



Yield: 33.6 %

^{1}H NMR: (500 MHz, D_2O) δ 7.21 (t, *J* = 9.0 Hz, 2H), 7.04 (t, *J* = 9.0 Hz, 2H), 4.37 (dd, *J* = 8.5, 5.0 Hz 1H) 3.12 (dd, *J* = 14.0, 5.0 Hz, 1H), 2.86 (dd, *J* = 14.0, 8.5 Hz, 1H) and 1.89 (s, 3H).

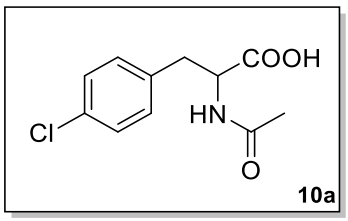
^{13}C NMR: (126 MHz, D_2O) δ 178.09, 173.24, 162.47, 160.55, 133.59, 133.56, 130.77, 130.70, 115.10, 114.93, 50.44, 36.91, 21.80.

^{19}F NMR: (500 MHz, D_2O) δ -117.49 (m, 1F).

TOF-ESI MS: calcd. for $C_9H_{11}ClFNO_2$ 224.1 $[M - H]^-$, found 224

Chiral HPLC: r R.T. = 9.01 min, *ee* > 99.9%

Commercially avaiable



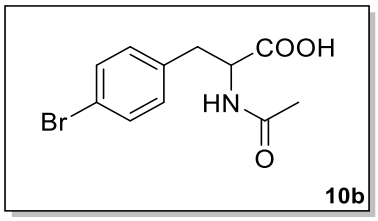
Yield: 62.9 %

^{1}H NMR: (500 MHz, D_2O) δ 7.34 (d, *J* = 8.0 Hz, 2H), 7.22 (d, , *J* = 8.0 Hz, 2H), 4.40 (dd, *J* = 9.0, 5.0f Hz 1H) 3.15 (dd, *J* = 14.0, 5.0 Hz, 1H), 2.88 (dd, *J* = 13.5, 8.5 Hz, 1H) and 1.90 (s, 3H).

^{13}C NMR: (126 MHz, D_2O) δ 177.99, 173.24, 136.48, 131.81, 130.68, 128.34, 56.25, 37.11 and 21.81.

TOF-ESI MS: calcd. for $C_{11}H_{12}O_3N_1Cl$ 242.0591 $[M + H]^+$, found 242.0581

Chiral HPLC: R.T. = 4.69 min (L) and 5.18 min (D), racemate



Yield: 85.3 %

^{1}H NMR: (500 MHz, D_2O) δ 7.40 (d, *J* = 8.0 Hz, 2H), 7.21 (d, *J* = 8.0 Hz, 2H), 4.40 (dd, *J* = 9.0, 4.9 Hz, 1H), 3.13 (dd, *J* = 13.9, 4.8 Hz, 1H), 2.85 (dd, *J* = 14.0, 9.0 Hz, 1H), 1.90 (d, *J* = 5.2 Hz, 4H).

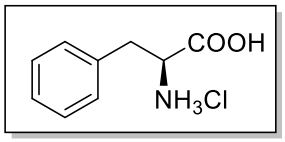
^{13}C NMR: (126 MHz, D_2O) δ 177.97, 173.23, 136.99, 131.32, 131.12, 119.93, 56.20, 37.17, 21.84.

TOF-ESI MS: calcd. for $C_{11}H_{12}O_3N_1Br$ 284.0074 $[M + H]^+$, found 284.0073

Chiral HPLC: R.T. = 4.68 min (L) and 5.22 min (D), racemate

Characterization of NAAAR/Acylase DKR scale-up reaction products

L-Phenylalanine·HCl



Yield: 88.8%

^{1}H NMR: (500 MHz, D_2O) δ 7.48 (m, 5H), 4.34 (t, *J* = 9.0 Hz, 1H), 3.38 (dd, *J* = 14.5, 7.5 Hz, 1H) and 3.25 (dd, *J* = 14.5, 7.5 Hz, 1H).

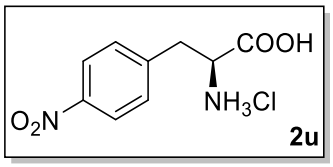
^{13}C NMR: (126 MHz, D_2O) δ 171.19, 134.16, 129.43, 129.24, 128.01, 54.47 and 35.72

TOF-ESI MS: calcd. for $C_9H_{12}ClNO_2$ 202.06 $[M + H]^+$, found 202.1

Chiral HPLC: R.T. = 9.20 min, *ee* > 99.9%

Commercially available.

L-4-NO_2-Phenylalanine·HCl (2u)



Yield: 61.9 %

^{1}H NMR: (500 MHz, D_2O) δ 8.10 (d, 2H), 7.38 (d, 2H), 3.48 (t, *J* = 9.0 Hz, 1H) 3.00 (dd, *J* = 13.5, 8.0 Hz, 1H) and 2.89 (dd, *J* = 13.5, 8.0 Hz, 1H).

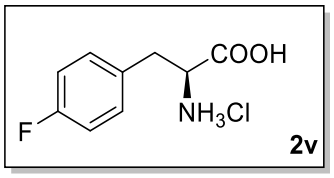
^{13}C NMR: (126 MHz, D_2O) δ 181.76, 146.96, 146.35, 130.29, 123.66 57.36 and 40.86.

TOF-ESI MS: calcd. for $C_9H_{11}ClN_2O_4$ 247.04 $[M + H]^+$, found 247.1

Chiral HPLC: R.T. = 9.83 min, *ee* > 99.9%

Commercially available.

L-4-F-Phenylalanine·HCl (2v)



Yield: 84.5 %

^{1}H NMR: (500 MHz, D_2O) δ 7 7.14 (m, 2H), 6.99 (t, *J* = 8.0 Hz, 2H), 3.37 (t, *J* = 9.0 Hz, 1H) 2.84 (dd, *J* = 13.5, 5.5 Hz, 1H) and 2.73 (dd, *J* = 13.5, 5.5 Hz, 1H).

^{13}C NMR: (126 MHz, D_2O) δ 171.79, 163.19, 161.25, 134.12, 134.07, 130.98, 130.92, 115.20, 115.03, 57.72and 40.03.

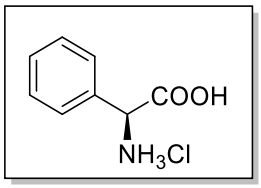
^{19}F NMR: (500 MHz, D_2O) δ -117.49 (m, 1F).

TOF-ESI MS: calcd. for $C_9H_{11}ClFNO_2$ 220.05 $[M + H]^+$, found 220.1

Chiral HPLC: r R.T. = 9.01 min, *ee* > 99.9%

Commercially available.

L-Phenylglycine·HCl



Yield: 50.2%

^{1}H NMR: (500 MHz, D_2O) δ 7.30 (m, 5H) and 4.27 (s, 1H).

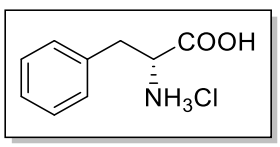
^{13}C NMR: (126 MHz, D_2O) δ 181.14, 142.21, 128.85, 127.54, 126.84 and 60.50

TOF-ESI MS: calcd. for $C_8H_{10}ClNO_2$ 188.04 $[M + H]^+$, found 188.0

Chiral HPLC: R.T. = 8.77 min, *ee* > 99.9%

Commercially available.

D-Phenylalanine·HCl



Yield: 74.0%

^{1}H NMR: (500 MHz, D_2O) δ 7.39 (t, 2H), 7.30 (m, 3H), 3.51 (t, *J* = 9.0 Hz, 1H), 3.00 (dd, *J* = 13.5, 5.5, 1H) and 2.84 (dd, *J* = 13.5, 7.5, 1H).

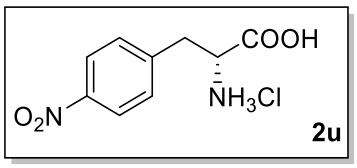
^{13}C NMR: (126 MHz, D_2O) δ 182.47, 138.33, 129.46, 128.65, 126.71, 57.49 and 40.87

TOF-ESI MS: calcd. for $C_9H_{12}ClNO_2$ 202.06 $[M + H]^+$, found 202.1

Chiral HPLC R.T. = 10.31 min, *ee* = 79.4%

Commercially available.

D-4-NO_2-Phenylalanine·HCl (2u)



Yield: 91.7%

^{1}H NMR: (500 MHz, D_2O) δ 8.27 (d, *J* = 9.1 Hz, 2H), 7.58 (d, *J* = 9.1 Hz, 2H), 4.36 (t, *J* = 9.0 Hz, 1H) 3.48 (dd, *J* = 14.5, 6, 1H) and 3.38 (dd, *J* = 14.5, 7.5, 1H).

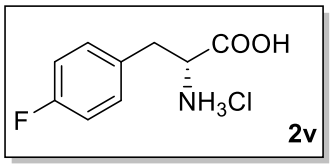
^{13}C NMR: (126 MHz, D_2O) δ 171.16, 147.23, 142.50, 136.22, 130.52, 124.17, 122.99, 59.35 and 35.62.

TOF-ESI MS: calcd. for $C_9H_{11}ClN_2O_4$ 247.04 $[M + H]^+$, found 247.1

Chiral HPLC R.T. = 10.86 min, *ee* = 27.7%

Commercially available.

D-4-F-Phenylalanine·HCl (2v)



Yield: 74.1%

^{1}H NMR: (500 MHz, D_2O) δ 7.36 (m, 2H), 7.19 (m, 2H), 4.32 (t, *J* = 9.0 Hz, 1H), 3.36 (dd, *J* = 14.5, 5.5, 1H), 3.25z (dd, *J* = 14.5, 7.5, 1H)

^{13}C NMR: (126 MHz, D_2O) δ 171.79, 163.19, 161.25, 131.23, 131.17, 130.00, 129.98, 115.97, 115.79, 59.37 and 34.95

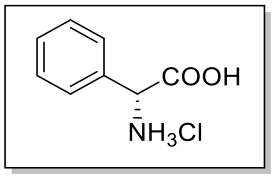
^{19}F NMR: (500 MHz, D_2O) δ -115.46 (m, 1F)

TOF-ESI MS: calcd. for $C_9H_{11}ClFNO_2$ 220.05 $[M + H]^+$, found 220.1

Chiral HPLC R.T. = 9.99 min, *ee* = 28.5%

Commercially available.

D-Phenylglycine·HCl



Yield: 90.6%

^{1}H NMR: (500 MHz, D_2O) δ 7.39 (m, 5H) and 4.36 (s, 1H)

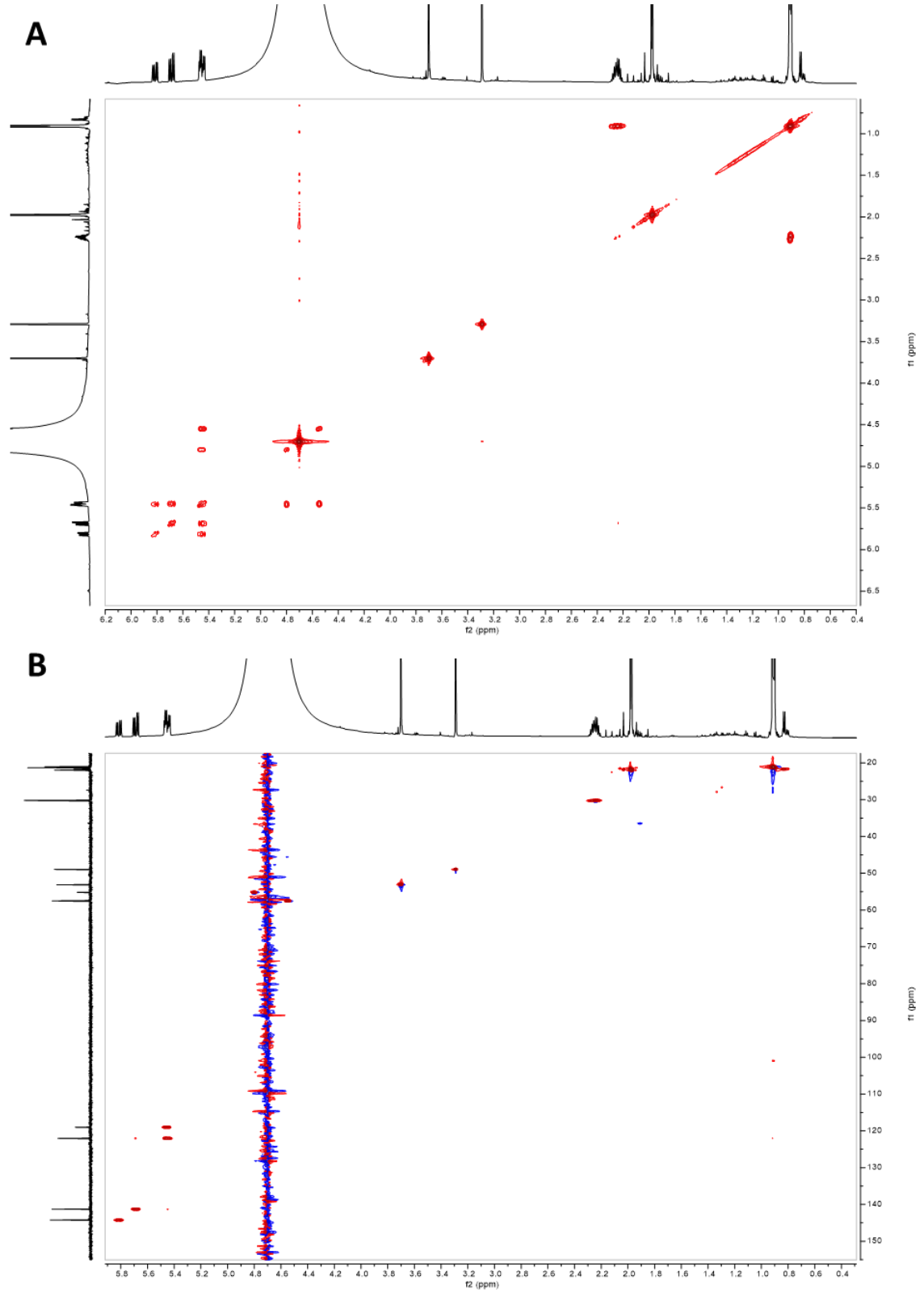
^{13}C NMR: (126 MHz, D_2O) δ 181.14, 142.17, 128.85, 127.54, 126.84 and 60.59

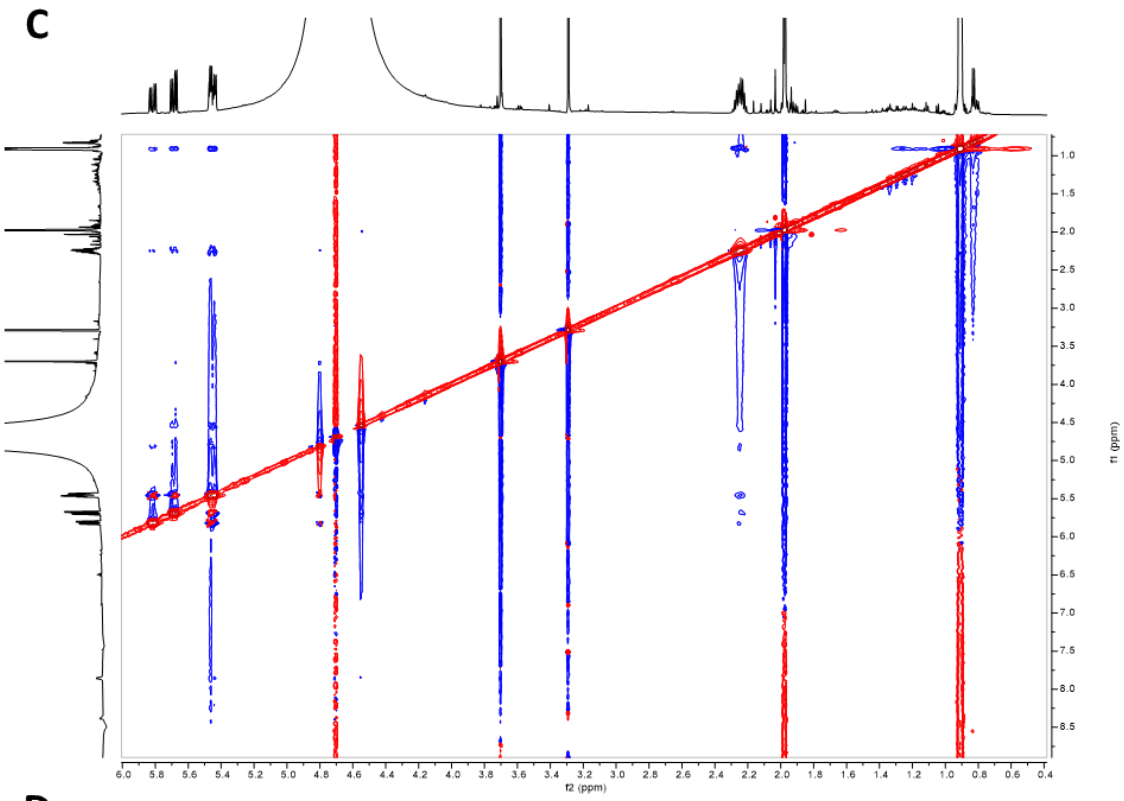
TOF-ESI MS: calcd. for $C_8H_{10}ClNO_2$ 188.04 $[M + H]^+$, found 188.0

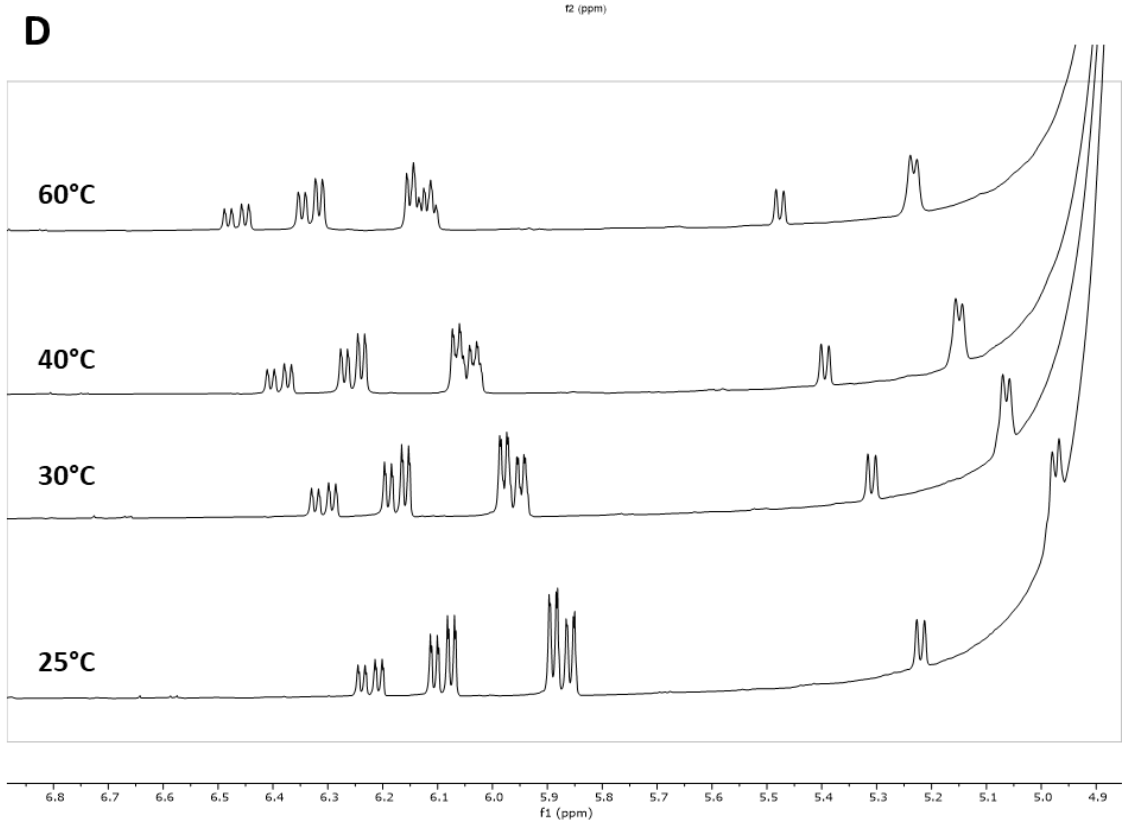
Chiral HPLC R.T. = 15.56 min, *ee* = 81.4%

Commercially available.

Appendix 22 Full NMR charachterization of compound **1l**. **A)** COSY NMR. **B)** HSQC NMR. **C)** NOESY NMR. **D)** Overlay of **1l** ^{1}H-NMR taken at various temperatures (25, 30, 40 and 60°C).




C
f2 (ppm)
f1 (ppm)


D
60°C
40°C
30°C
25°C
f1 (ppm)